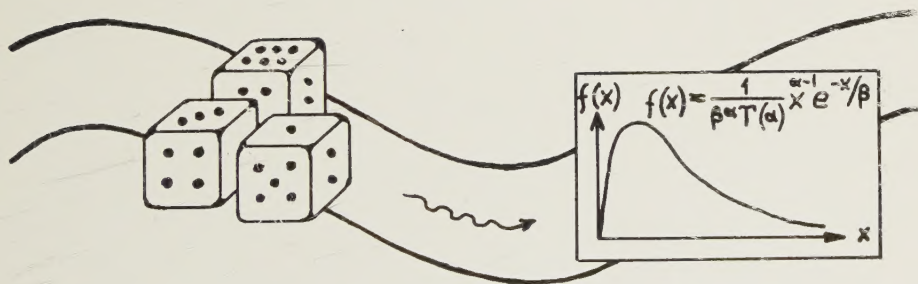
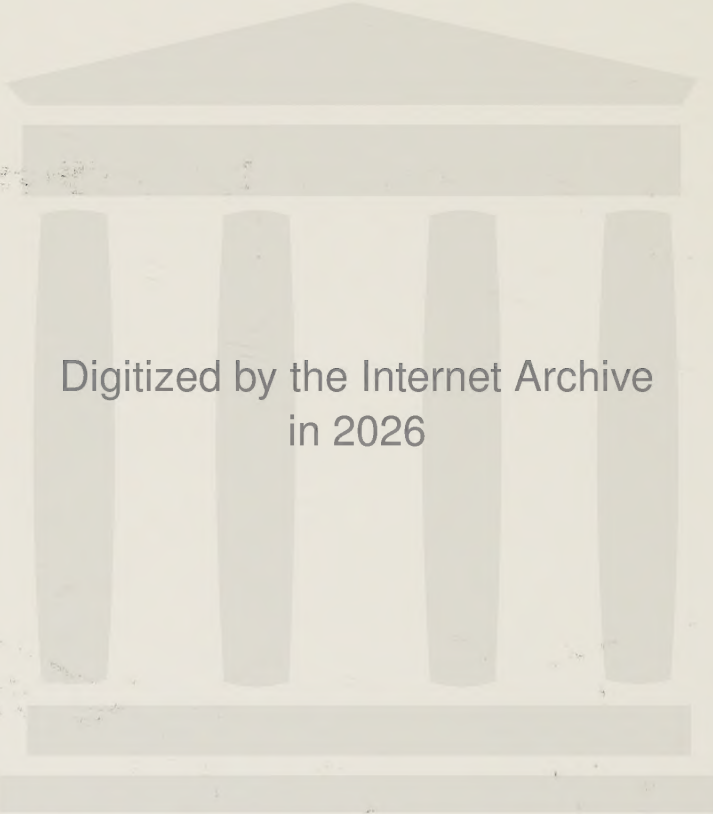


STOCHASTIC MODELLING OF MONTHLY RIVER RUNOFF

by

Lars Gottschalk





Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41552412_0011

INSTITUTIONEN FÖR TEKNISK VATTENRESURSLÄRA
LUNDS TEKNISKA HÖGSKOLA / LUNDS UNIVERSITET

DEPARTMENT OF WATER RESOURCES ENGINEERING
LUND INSTITUTE OF TECHNOLOGY / UNIVERSITY OF LUND

STOCHASTIC MODELLING OF MONTHLY RIVER RUNOFF
A study based on multivariate distribution functions
by
Lars Gottschalk





CONTENT

	page
Forward	1
Summary	2
Summary in Russian	5
List of tables	8
List of illustrations	10
List of general symbols	12
List of special symbols	14
1. Introduction	17
2. Natural river runoff as a random process	26
3. Marginal distribution functions for monthly average runoff	31
3.1 Introduction	31
3.2 Empirical distributions	33
3.2.1 Empirical moments	34
3.2.2 Order statistics	35
3.3 Regional variations in empirical distribution of the module of average monthly runoff	38
3.3.1 Regional variations in empirical moments	45
3.3.1.1 Polar Rivers	45
3.3.1.2 Northern Mountain Rivers	53
3.3.1.3 Central Mountain Rivers	55
3.3.1.4 Mountain and Forest Rivers in the South of Norrland and North of Svealand	57
3.3.1.5 Forest and Coastal Rivers in Norrland	60
3.3.1.6 Rivers of the South of Sweden	62
3.3.1.7 Rivers from the large Swedish Lakes	69
3.3.2 General conclusions	72
3.4 Analytical distribution functions	76
3.4.1 Transformations of the normal distribution	79
3.4.2 Empirical transformation functions of the unit normal distribution	82
3.4.3 Analytical transformation functions of the unit normal distribution	88
3.4.4 Transformations of the gamma distribution	89
3.4.5 Parameter estimation	91
3.4.5.1 The method of maximum likelihood	93
3.4.5.2 The method of moments	99

	page
3.4.5.3 The graphical method	105
3.4.5.4 Conclusions	110
3.5 Test of marginal distribution functions	118
4. Multivariate distribution functions	123
4.1 Scatter diagrams	125
4.2 The bivariate normal distribution	125
4.3 The bivariate gamma distribution	129
4.4 The multivariate normal distribution	132
4.5 Parameter estimation	134
4.5.1 The coefficient of correlation	135
4.5.2 Regional variations in the coefficient of correlation	137
4.5.3 Parameters of the bivariate Γ -distribution	148
5. Finite difference equation for the runoff process	151
5.1 Scalar process	151
5.2 Vector process	156
5.3 The order of the Markov process	163
5.3.1 Test of normality of the residual	163
5.3.2 Test of independence of the residual	164
5.3.3 Test of variance of the residual	165
5.3.4 Test of the regression coefficient	167
5.3.5 Other tests	168
5.3.6 Conclusions about the order	169
5.4 Generation scheme for the bivariate gamma distribution	170
5.5 The use of the stochastic finite difference equation for river runoff	171
6. Properties of the models	175
6.1 Moments of monthly runoff	177
6.2 Moments of annual runoff	185
6.3 Other properties	191
6.4 General conclusions	191
7. Conclusions	196
References	197

FOREWORD

The present publication deals with the problem of stochastic modeling of monthly river runoff for water management calculations. This theme was chosen after the suggestion of professor Nikolaj Archilovich Kartvelishvili during my stay as an apprentice at the Geographical faculty of the Lomonosov Moscow State University in 1972-73. The theoretical questions of this problem was studied under the excellent and most inspiring supervision of professor Nikolaj Archilovich Kartvelishvili. Valuable knowledge was obtained during lectures and discussions at the Geographical faculty, the Laboratory of Water Problems and Higher Hydrological Courses.

The work was continued and completed at the department of Water Resources Engineering, University of Lund, under the leadership of professor Gunnar Lindh. It was partly economically supported by the Swedish IHD-committee.

I want to express my deepest gratitude to professor Nikolaj Archilovich Kartvelishvili and professor Gunnar Lindh for their guidance and support in my studies.

Valuable technical work and redaction of the manuscript was done by my wife Irina Krasovskaja; the laborious task of manuscript typing was performed by Miss Kerstin Krahner; illustrations and drawings were skillfully done by Mrs Sonja Persson, Malmö Kartritningstjänst and I want to thank them for their help.

I am grateful to my colleagues at the Geographical faculty of the Lomonosov Moscow State University and of the department of Water Resources Engineering, University of Lund for their encouraging me and their remarks on the work.

Lund, March 1975

Lars Gottschalk

SUMMARY

Water management calculations assume that the full probability characteristics of the natural runoff process are known. The present work, with its assumptions about the distribution functions for monthly river runoff, gives the necessary characterization of this process for such calculations on monthly basis.

It is assumed that the river runoff process for monthly average runoff values, can be approximated by a Markov process, and is harmonic, with the period of one astronomic year, and ergodic. The process is thus described by twelve distribution functions:

$$F_n(q_t, t; q_{t-1}, t-1; \dots; q_{t-n}, t-n), t = 1, \dots, 12$$

where n is the order of the Markov process and q_t , in general case, is a vector representing runoff at several sites in a river or neighbouring rivers.

A fundamental problem when dealing with multivariate distributions, as the one given above, is to describe the relation between the multivariate distribution and the corresponding marginal distributions. If the marginal variables considered, are distributed normally, we can assume that they have a joint multivariate normal distribution. For other marginal distributions bivariate distributions can be given as series of certain sets of orthonormal functions and canonical correlation coefficients. In case of gamma distribution generalized orthonormal Laguerre functions are used.

The regional and seasonal variations of the empirical distributions of monthly runoff have been studied by means of the average values and the coefficients of variation and skewness. A clear distinction can be made between the northern and southern parts of Sweden. For the northern rivers five different periods within the year are identified. Analytical distribution functions that describe well the observational data are the two-parameter log-normal and gamma distributions and some other transformations of the unit normal distribution. The southern rivers have three different periods. The physiographic differences between the northern and southern regions are not reflected by the type of analytical distribution function applied. The best

results for the southern region are also reached by the same analytical distributions as above.

To find suitable analytical expressions for transformations of the normal distribution empirical transformation functions have been studied. These functions, of course, to a great extent are described by the coefficients of variation and skewness, and thus show the same regional and seasonal differences as these coefficients.

Different methods for parameter estimation for transformations of the form $u(x) = \sum_{i=1}^n \alpha_i \phi_i(x)$ of the unit normal distribution have been discussed. For two parameter transformations ($n = 2$) differences in the estimates derived by the different methods are not very great. For higher values of n the discrepancies can be significant.

Parameters of multivariate normal distribution - the correlation coefficients, and of bivariate gamma distribution - the canonical correlation coefficients have been calculated for monthly runoff values. The correlation coefficient shows an annual periodic variation in the same way as the empirical marginal moments. This variation is most apparent for the northern groups of rivers. Climatic runoff forming factors seem to explain most of the variations in the correlation coefficient. The correlograms for both groups of rivers are functions of time. The runoff process is thus nonstationary. The canonical correlation coefficients of the bivariate gamma distribution derived by the method of moments show low accuracy and in many cases the series expansion do not converge.

Finite difference equation of the multivariate normal distribution for scalar and vector runoff processes are given. Such equations are better for simulation purposes than the direct use of the multivariate distribution functions. A proper order of the Markov process can be found by applying statistical tests on the coefficients of the finite difference equations. Different tests give contradictory results. Best results are achieved with an F-test on the variance of the residual. For both the northern and the southern groups of rivers a second order Markov process seems to be proper. Exception should be made for outflows from the large Swedish lakes, where still higher orders are recommended.

Properties of the found models for monthly runoff process have been examined, with respect to how well monthly and annual moments are preserved. Marginal and mixed moments of monthly river runoff are preserved only in the case, when the method of moments is used. Significant differences in the theoretical moments of the models and the observed ones were noticed when other methods of estimation are used. Which particular properties of the model are important and should be preserved, is determined by the specific practical application of the model. To preserve annual moments of the runoff process are a more difficult task. They depend on the monthly marginal distributions as well as the correlation structure within the year. The moments vary with the definition of the beginning of the year, due to the instationarity of the runoff process. Other criteria for acceptability of a runoff model have been discussed.

РЕЗЮМЕ

Водохозяйственные расчеты предполагают знание полной вероятностной характеристики природы стокового процесса. Настоящая работа, основываясь на предположениях о функциях распределения речного стока, дает характеристику стокового процесса, необходимую для подобных расчетов.

Было принято, что для среднемесячных величин процесс речного стока может быть аппроксимирован Марковским процессом и является периодическим, с периодом в один год, и эргодичным. Таким образом, этот процесс будет описываться двенадцатью функциями распределения:

$$F_n(q_t, t; q_{t-1}, t-1; \dots; q_{t-n}, t-n), t = 1, \dots, 12$$

где n - порядок Марковского процесса,
а q_t - в общем случае, вектор, представляющий сток в ряде пунктов на одной реке или на соседних реках.

При использовании многомерных распределений, таких как приведенное выше, фундаментальной задачей является описание зависимости между многомерным распределением и соответствующими маргинальными распределениями. Если рассматриваемые маргинальные величины распределены нормально, то мы можем предположить, что они имеют совместное многомерное нормальное распределение. Для иных маргинальных распределений могут быть даны двумерные распределения в виде рядов определенных ортонормальных функций и канонических коэффициентов корреляции. В случае гамма-распределения используются обобщенные ортонормальные функции Лагерра.

При помощи среднемесячных величин стока, коэффициентов вариации и асимметрии изучались региональные и сезонные колебания эмпирических распределений среднемесячного стока. Было замечено четкое различие между северной и южной Швецией. Для северных рек было выделено пять различных периодов в течение года. Аналитическими распределениями, хорошо описывающими данные наблюдений,

являются логнормальное распределение и некоторые другие трансформации нормального распределения. Южные реки имеют три различных периода. Тип аналитической функции распределения не отражает физико-географические различия между северным и южным районами. Аналитические распределения, упомянутые выше, дают наилучшие результаты также для южного района.

С целью найти наиболее подходящие аналитические выражения для трансформаций нормального распределения были изучены эмпирические трансформирующие функции. Эти функции, конечно, в значительной степени описываются коэффициентами вариации и асимметрии и поэтому имеют такие же региональные и сезонные различия, как и эти коэффициенты.

Рассматривались различные методы оценки параметров трансформации вида:

$$u(x) = \sum_{j=1}^n \alpha_j \phi_j(x),$$

единичного нормального распределения. Для трансформаций с двумя параметрами ($n = 2$) различия в оценках, полученных методом наибольшего правдоподобия, не очень велики. Для более высоких величин "n" расхождения могут быть значительными.

Были подсчитаны параметры многомерных нормальных распределений - корреляционные коэффициенты и параметры двумерного гамма-распределения - канонические коэффициенты корреляции. Коэффициент корреляции имеет ежегодные периодические колебания, аналогично эмпирическим маргинальным моментам. Эти колебания наиболее очевидны для северной группы рек. Климатические стокообразующие факторы объясняют, по всей вероятности, большую часть колебаний корреляционного коэффициента ежемесячного стока. Коррелограммы для обеих групп рек являются функциями времени. Процесс стока, таким образом, - нестационарный.

Канонические коэффициенты корреляции двумерного гамма-распределения, полученные методом моментов, имеют низкую точность, и во многих случаях продленные ряды не сходятся. Приводятся уравнения конечных разностей для многомерного нормального распределения для скалярного и векторного процессов стока. Подобные уравнения лучше отвечают целям моделирования, чем прямое использование многомерных функций

распределения. Нужный порядок Марковского процесса может быть найден посредством применения статистических тестов для коэффициентов уравнений конечных разностей. Различные тесты дают противоречивые результаты. Лучшие результаты получаются при использовании F-теста для остаточной дисперсии. Для обеих - северной и южной групп рек подходит, очевидно, второй порядок Марковского процесса. Исключение следует сделать лишь для стока из больших шведских озер, для которого рекомендуются еще более высокие порядки.

Для месячного процесса стока были проанализированы свойства найденных моделей, с учетом того, насколько хорошо сохраняются ежемесячные и годовые моменты. Маргинальные и смешанные моменты месячного речного стока сохраняются только в том случае, когда используется метод моментов. При использовании других методов оценки отмечаются значительные различия между теоретическими моментами модели и наблюдаемыми. Специфическое практическое использование модели определяет, какие, именно, ее свойства представляют важность и должны быть сохранены. Более трудной задачей является сохранение годовых моментов стокового процесса. Они зависят как от месячных маргинальных распределений, так и от корреляционной структуры внутри года. Годовые моменты изменяются в зависимости от определения начала года вследствие нестационарности процесса речного стока. Рассматриваются также некоторые другие критерии приемлимости модели речного стока.

LIST OF TABLES

- Table 3.1 Name of observation points, name of the river and period of observation of data used.
- Table 3.2-25 Monthly means and coefficients of variation and skewness of river runoff for different hydrological regions.
- Table 3.26 Range of variation of monthly means and coefficients of variation and skewness for different hydrological regions.
- Table 3.27 Example of calculated transformed β -values
 $u_i^1 = \Phi^{-1}(p_i)$ for $N = 41$.
- Table 3.28-33 Examples of estimated parameters with maximum likelihood, moment and graphical methods for different transformations of the unit normal distribution.
- Table 3.34-36 Result of χ^2 -test of different marginal distributions for monthly river runoff.
- Table 4.1 Space correlation matrices for different months for the observation points 1-959 Abiskojokk, 1-3 Jukkasjärvi and 1-589 Kallio.
- Table 4.2 Ditto for 101-224 Johansfors, 103-974 Kila and 105-227 Åsbro.
- Table 4.3 Monthly parameters of the bivariate gamma distribution for 86-1008 N Forsmöller.
- Table 5.1 Number of rejections of hypothesis that the residuals of stochastic finite difference equations are unit normally distributed.
- Table 5.2 Residual standard deviation for different months and for different orders of Markov process for 28-444 Hällbacken.

Table 5.3	Examples of residual standard deviation of vector models for different orders of Markov process for January.
Table 6.1	Observed and calculated monthly expected values for 96-1008 N Forsmöllen.
Table 6.2	Observed and calculated monthly standard deviation for 96-1008 N Forsmöllen.
Table 6.3	Observed and calculated monthly skew coefficients for 99-1008 N Forsmöllen.
Table 6.4	Observed and calculated correlation coefficients of original and logarithm monthly runoff values.
Table 6.5	Range of standard deviation calculated from ten simulated series with a first order Markov model with normal and lognormal distributions, respectively for 96-1008 N Forsmöllen.
Table 6.6	Annual variation coefficients as a function of the beginning month of the year for different observation points.
Table 6.7	Ditto for annual skew coefficients.
Table 6.8	Ditto for log one correlation coefficient.
Table 6.9	Range of annual correlation coefficients calculated from ten simulated series of different models for 1-589 Kallio.

LIST OF ILLUSTRATIONS

- Fig 1.1 Distribution functions of annual regulated reservoir volume, analytically derived from the distribution of annual river runoff.
- Fig 1.2 Marginal distributions of monthly runoff of the lake Bolmen and calculated marginal distribution with the Monte-Carlo method of the levels of lake Bolmen.
- Fig 3.1 Map over Swedish regions with equal runoff forming factors (after Kuprijanov, 1960).
- Fig 3.2 Map over Swedish river regimes (after Melin, 1970).
- Fig 3.3 Map over Swedish hydrological regions (after Melin, 1970).
- Fig 3.4-5 Examples of observed runoff series from different hydrological regions.
- Fig 3.7 Map over used observation points.
- Fig 3.8-16 Graphs of means and coefficients of variation and skewness for different hydrological regions.
- Fig 3.17-18 Examples of plotted empirical distributions of river runoff for some months.
- Fig 3.19 Graphical illustration of transformation of the unit normal distribution.
- Fig 3.20-25 Empirical transformation functions of the unit normal distribution for monthly river runoff for different hydrological regions.
- Fig 3.26 Standard deviation of transformed β -values s_u as a function of the number of observations N .
- Fig 3.27 Examples of transformation curves of the unit normal distribution with parameters estimated with different methods.

- Fig 4.1 Scatter diagrams of module values of monthly runoff for February - March and April, May at 28-444 Hällbacken.
- Fig 4.2 Scatter diagrams of standardized and normalized module values of river runoff for November-December at 4-18 Mojärv.
- Fig 4.3 Bivariate normal distributions for data given in Fig 4.1.
- Fig 4.4 Bivariate log-normal distribution for data given in Fig 4.1.
- Fig 4.5 Annual variation of first to third order correlation coefficients of standardized and of normalized data for monthly river runoff at 1-589 Kallio.
- Fig 4.6 Ditto at 101-224 Johansfors.
- Fig 4.7 Annual variation of first order correlation coefficient for monthly river runoff at 67-154 Motala.
- Fig 4.8 Monthly correlograms for some different months for 28-444 Hällbacken.
- Fig 4.9 Ditto for 74-177 Järnforsen.
- Fig 5.1 Coefficients of first to fifth order stochastic finite difference equation for monthly river runoff for 75-855 Getebro.
- Fig 6.1-2 Examples of simulated series with different models.
- Fig 6.3 Observed and calculated, with confidence intervals, distributions of monthly outflow from the lake Bolmen.

LIST OF GENERAL SYMBOLS

a_k	estimate of k-th order moment of origin
$\{c_k\}$	parameter set in orthogonal series expansion
Cov	covariance
C_s	coefficient of skewness
C_v	coefficient of variation
$C_{v,i}^{cond}$	conditional coefficient of variation
D	determinant
D_{ij}, D_{ij}^{nk}	cofactors to the determinant D
D_u	determinant of correlation matrix
$E(x)$	expected value
$F(x)$	probability distribution function
$F(x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_n)$	probability distribution function for variables $x_1, x_2, \dots, x_n$ and with parameters $\theta_1, \theta_2, \dots, \theta_n$
$F^x(x)$	empirical probability distribution function
F^n, F_k^n	test quantities
$f(x)$	probability density function
$f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_n)$	probability density function for variables $x_1, x_2, \dots, x_n$ and with parameters $\theta_1, \theta_2, \dots, \theta_n$
H_s	parameter matrix
H_{jk}	elements of H_s
L_k	generalized Laguerre polynomial
m_k	estimate of k-th order central moment
$\bar{m}, \hat{m}_n$	mean
$\bar{m}, \hat{m}_n, \bar{x}$	estimate of mean
$m_{x_2 x_1}$	conditional mean
$P(X \leq x)$	probability function
$P(X \leq x Y = y)$	conditional probability function
p, p_i, p_i', p_i''	order statistic
Q, Q_i, q_i	monthly average runoff
$Q(t), Q_i^{st}(t), q_i^{st}(t)$	monthly average runoff as a function of time
r, r_{ij}, r_{ij}^{st}	estimate of correlation coefficient
R_i	canonical correlation coefficient
s^2, s_n^2, s_x^2	estimate of variance
t	time variable
t_{nk}	test quantity
$U, u, u(x)$	transformation function to unit normal distribution

Var	variance
X, x, X_i, x_i	module values of monthly average runoff
$X(t), x(t), X_i(t), x_i(t)$	module values as a function of time
x_i^i	i-th ordered value of observed sample of X
α_k	k-th order, theoretical moment
β_{ij}	elements of unitary matrix
$\Gamma(\gamma)$	gamma function
Λ_n^2	residual variance
μ_k	theoretical k-th order central moment
$\rho, \rho_{ij}, \rho_{ij}^{st}$	theoretical correlation coefficient
$\sigma_z^2, \sigma_n^2, \sigma_x^2$	theoretical variance
χ^2	test quantity
$\phi(x)$	unit normal probability distribution function
$\phi'(x)$	unit normal probability density function

LIST OF SPECIAL SYMBOLS

Chap 1

a	constant
$B(t)$	experiences for a water resources system as function of time
L_t	function
m	number
N	number
S	discrete random variable
$s(x)$	function of x
T	length of time period
$W_i(t)$	reservoir volume as a function of time
y	integration variable
α	regulated discharge from a reservoir
β	reservoir volume
v_k	discrete random variable
τ	time lag
$\phi(x)$	probability distribution function

Chap 2

$F_i^x(x)$	probability distribution function
$F_n(q_1, t_1; \dots; q_n, t_n)$	multivariate probability distribution function
	dependent on time
r	number
S	number
T	time period
t_i	point in time
z	integration variable
θ	time period
τ_i	time interval
$f_s(j), \phi_r$	function

Chap 3

a, a', a'', a_γ	parameter
$\hat{a}$	estimate of a
A^i, A_{ij}, A_{ijk}	parameter
b, b', b'', b_γ	parameter

$\hat{b}$	estimate of b
B, B^i	parameter
c	parameter
$\hat{c}$	estimate of c
d_i	constant
e_i	constant
G_i	function
g	function
L	likelihood function
l	log-likelihood function
N	number
α_j	parameter
β	parameter
$\hat{\beta}$	estimate of β
γ	parameter
$\hat{\gamma}$	estimate of γ
δ	parameter
ε	parameter
$\hat{\varepsilon}$	estimate of ε
C	parameter
λ, λ_i	parameter
v_i	number
ξ	parameter
$\phi_i(x)$	function of x
ψ	function

Chap 4

$G(x)$	probability distribution function
$H(y)$	probability distribution function
k	parameter
N	number
$x^{(i)}$	set of orthonormal polynomials
$y^{(i)}$	set of orthonormal polynomials
β, β_i	parameter
γ, γ_i	parameter
$\Delta_{i,k}$	correlation matrix
δ	parameter
λ	parameter

Chap 5

B, B_j	parameter
C	parameter
E	determinant
$E_{j \ k}$	minor to the determinant E
L_k	parameter matrix
M_k	parameter matrix
y_i	variable
α_i	parameter
β_i	parameter
γ_i	parameter
ϵ	variable
n	variable
λ_i	eigenvalue
μ	matrix
μ_{ij}	element of μ
ν_{ij}	parameter
ξ_i	variable
$\psi_i(x)$	function of x

Chap 6

a	parameter
b	parameter
g_i^j	polynomial
h_i^j	polynomial
β_j	parameter
γ_j	parameter

1. INTRODUCTION

The problems of optimal design and operation of different water management systems, risk for failure and its scale are of vital importance in modern water resources planning. In order to treat these problems properly, i.e. bearing in mind the random character of the input to the systems - river runoff, it is necessary to describe it by its probability distribution functions. Probability distribution functions should be understood in a broad sense, which means that conditional distributions as well as multivariate ones can be involved. The type of distribution used should satisfy the demands of every particular problem and depends on whether design or operation rules are of interest.

The probability distributions of river runoff are in fact a description of the risks of the occurrence of different flows. The problem is to evaluate the risks of certain design and operation rules, taking the probability characteristics of the natural flow as a basis. Let us discuss a simple water resources system with one regulated reservoir. The calculations can be based on the water balance equation for this reservoir. We use annual runoff values and assume that they are independent. A theoretical treatment of the problem leads to the equation (Kartvelishvili, 1967):

$$\Phi(x) = F(x - \beta + \alpha) + \int_0^{\beta} \Phi(y) f(x - y + \alpha) dy \quad (1.1)$$

where

x is the module of annual flow

α annual discharge from the reservoir

β reservoir volume

$F(x)$ the probability distribution function
of natural flow

$f(x)$ the probability density function,
corresponding to $F(x)$

$\Phi(x)$ the probability distribution function of
regulated reservoir volume

Analytical solutions to this equation for the case of gamma distributed natural flow is given by Kiktenko and Oshirov (1964a). With the help of numerical quadrature the equation can be easily solved numerically (Kiktenko and Oshirov 1964b, 1965, Kartvelishvili 1967). An example for this is given in Fig 1.1, where the equation (1.1) has been solved for some different values of α and β for the given distribution function of the natural flow. From the derived distribution curves we can estimate, for instance, the probability of the reservoir being empty and full. For more complex cases of dependent inflow and a multisite reservoir system, generalizations of equation (1.1) has been given by Kartvelishvili and Korganova (1973). From these, we can derive the following integral equation system for dependent monthly flow

$$\begin{aligned} \Phi_k(x_k) = & \int_0^{x_k} \int_0^{\infty} f_k \{x_k - s(x_{k-1}) + \alpha; x_{k-1} - \beta + \alpha\} dx_k dx_{k-1} \\ & - \int_0^{\infty} \Phi_{k-2}(x_{k-2}) \frac{\partial}{\partial x_{k-2}} \left[\int_0^{x_k} \int_0^{\infty} f_k \{x_k - s(x_{k-1}) + \alpha; \right. \\ & \left. x_{k-1} - s(x_{k-2}) + \alpha\} dx_k dx_{k-1} \right] dx_{k-2}, \quad k = 1, \dots, 12 \end{aligned} \quad (1.2)$$

where the notations are the same as in eq (1.1) except the index of the month k and the function $s(x_k)$, which is defined by

$$s(x_k) = \begin{cases} 0 & x_k < 0 \\ x_k & 0 < x_k < \beta \\ \beta & x_k > \beta \end{cases}$$

Further, it should be noted that we are dealing here with bivariatic distribution functions, as the natural monthly flows are dependent.

The numerical solution of this equation system is very laborious. A more straightforward approach can be solution of this equation by using synthetic realizations at the natural runoff, on which we perform an ordinary regulation calculation. The synthetic realisations are generated from the set of bivariate distribution functions $f_k(x_k, x_{k-1})$, $k = 1, \dots, 12$ for the natural flow. Then we perform a statistical analysis on the results of the calculations to derive the distribution functions $\Phi_k(x_k)$, $k = 1, \dots, 12$. This way of

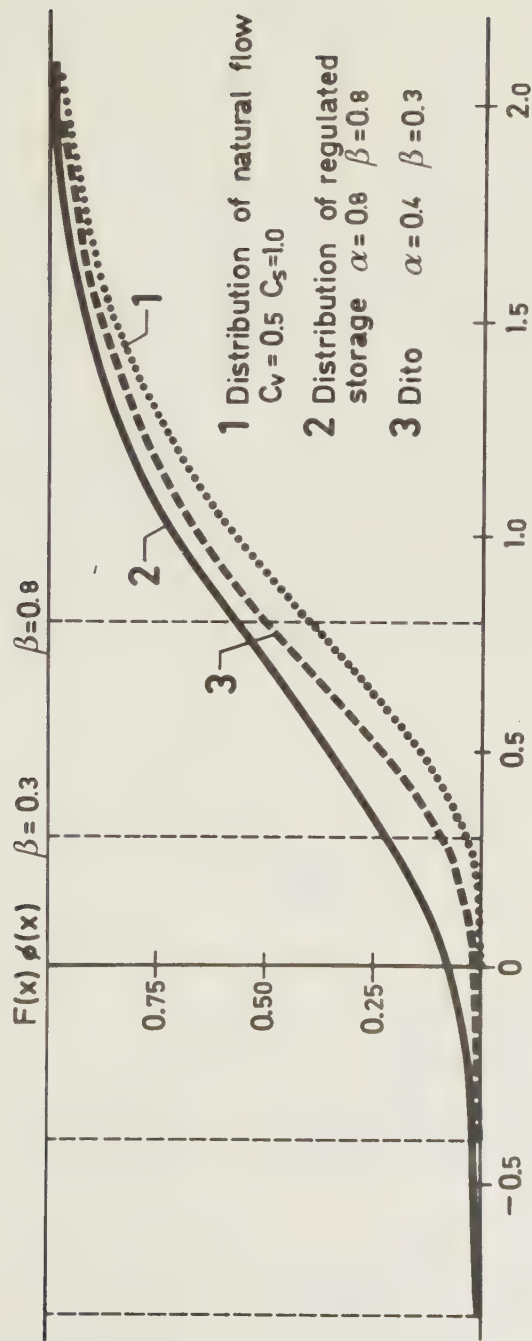


Fig 1.1 Distribution functions of annual regulated reservoir volume, analytically derived from the distribution of annual river runoff.

solving the equation system is usually called the Monte-Carlo method. The weak point of this method is generally that it converges to its solution very slowly, i.e. the synthetic realization must be very long.

An example of such calculation is shown in Fig (1.2), where the distribution functions of natural flow for each month of the lake Bolmen in the river Lagan are given. Only the marginal distributions are shown in the figure. The lake Bolmen is planned to supply the southwest part of Scania with drinking water and a regulation plan has been set up for this purpose. The regulation calculation was carried through on a realization of 1000 years of flow. The resulting distribution functions for each month are shown in Fig 1.2 together with the upper and lower limits for the regulation. The probability of failure of the regulation plan can be estimated now. Knowing the monthly distribution functions the probability of failure in a year, say $x_k < a$, where a is the given limit of the regulation plan, can be calculated in the following way. (Kartvelishvili and Korganova, 1973). We introduce the random variable v_k , that is defined by

$$v_k = 1 \text{ for } x_k \leq a$$

and

$$v_k = 0 \text{ for } x_k > a$$

Then the number S of months with failure in a year, is given by the sum

$$S = \sum_{i=k}^{k+11} v_i \quad (1.3)$$

S is also a random variable with the expected value:

$$\begin{aligned} E(S) &= \sum_{i=k}^{k+11} E(v_i) = \sum_{i=k}^{k+11} \{1 \cdot P(v_i = 1) + 0 \cdot P(v_i = 0)\} = \\ &= \sum_{i=k}^{k+11} \Phi_i(a) \end{aligned} \quad (1.4)$$

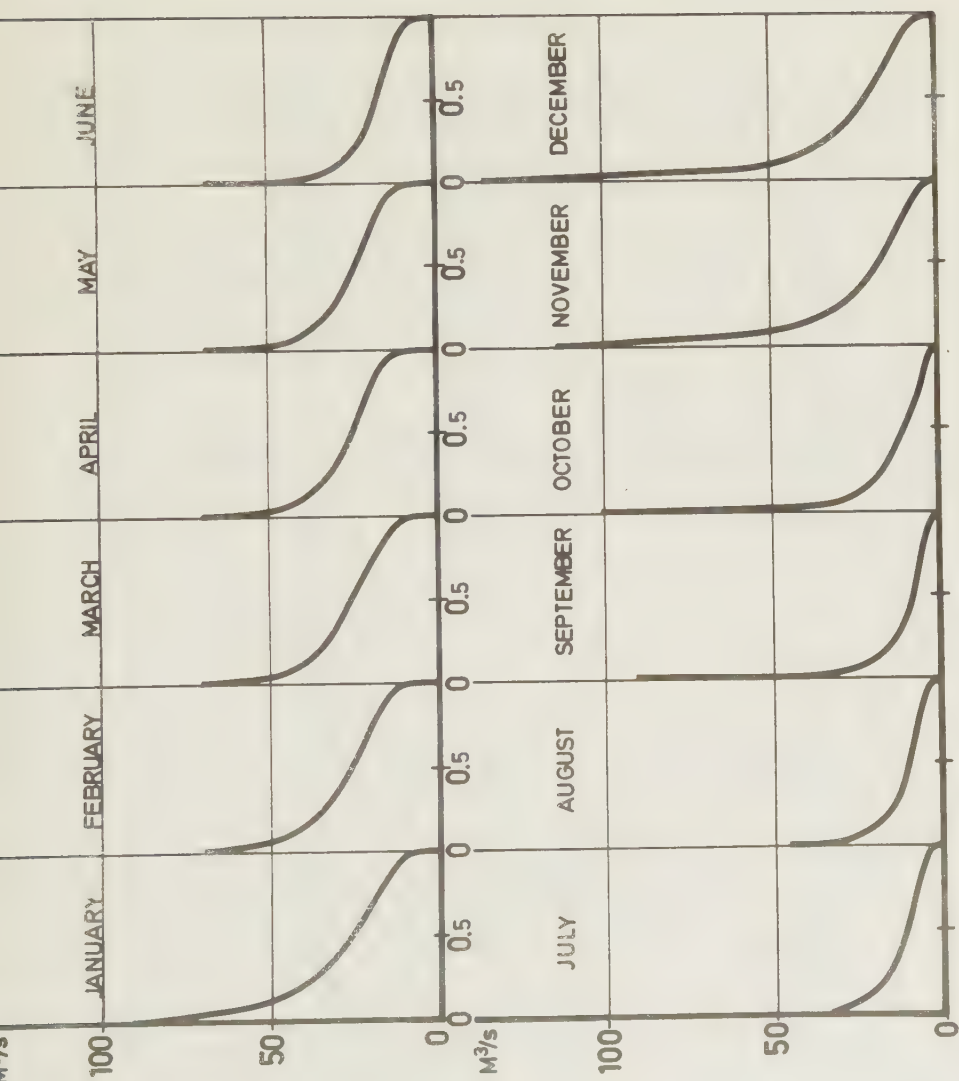


Fig 1.2 Monthly marginal distributions of river runoff of the lake Bolmen and calculated ones with the Monte-Carlo method of the levels of lake Bolmen.

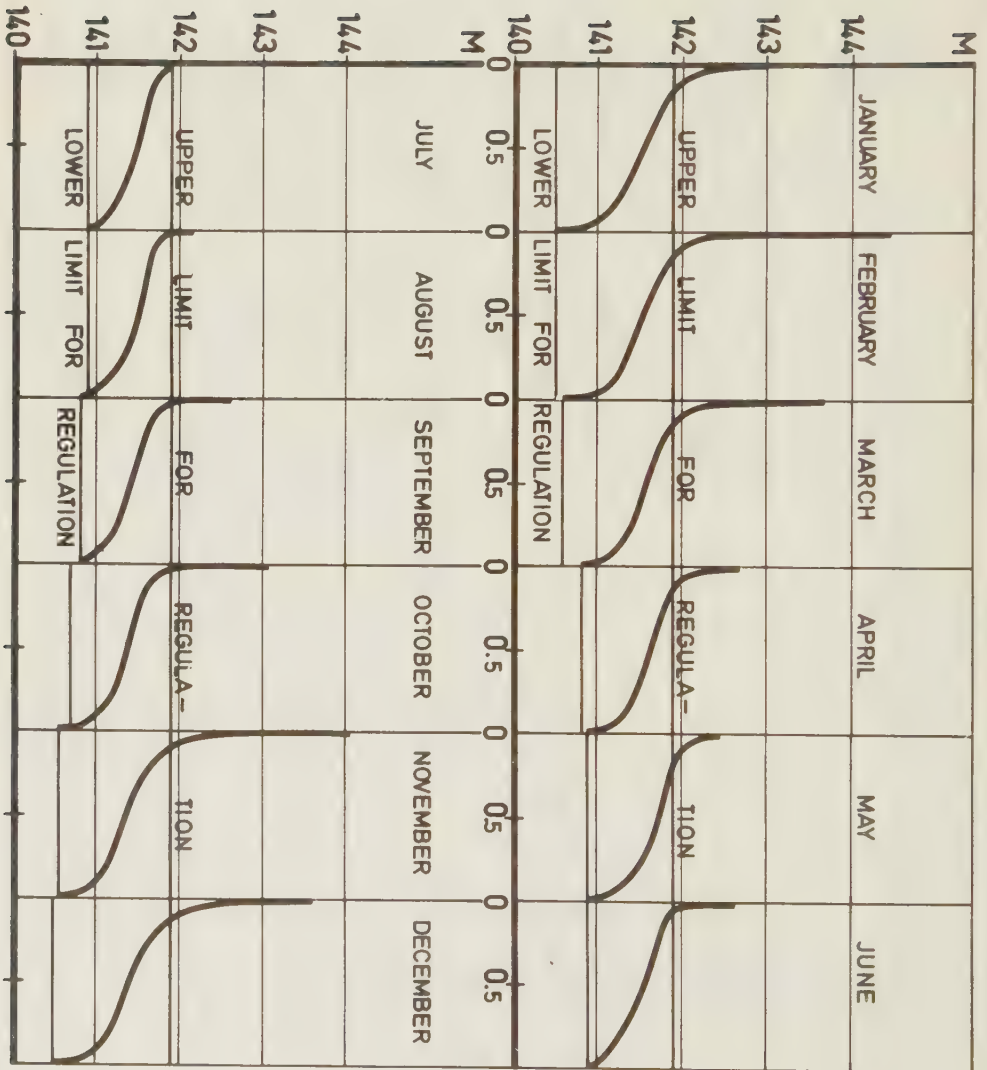


Fig 1.2 Continued

The probability of failure in a year can then be estimated by:

$$\frac{E(S)}{12} = \frac{1}{12} \sum_{i=k}^{k+11} \phi_i(a) \quad (1.5)$$

In the example given this probability is 0.033.

The probability of failure does not tell anything about the severity of it. The length of a period with low flow in a river or with low water level in a lake, and the total deficit of water during a low water period describe it better. These parameters are also of a random nature, as they are dependent on the river runoff, and therefore are characterized by their distribution functions. Principally they could be calculated directly from the known distribution functions of river runoff, but usually the Monte-Carlo method has to be used to derive them.

There exist several methods to find optimal design or optimal operation rules in water management, such as variational calculus, dynamic programming and Pontrjagin maximum principle. The optimum, due to the randomness of the input function to the system (the river runoff) is a random variable, and therefore is not well defined. To avoid this, the optimum of the expected value is usually taken as the optimal solution. This can be more or less advantageous, dependent on the character of the distribution function of the optimum.

The least difficult method for this kind of calculation is, therefore, the Monte-Carlo method, especially considering the stochastic nature of the problem. Let us consider, for instance, a power production system consisting of hydropower stations and thermal power stations. In the system we have m reservoirs. The inflow to the i -th reservoir at the time t is $Q_i(t)$ and the volume of it $W_i(t)$. The expenses can then be expressed (Kartvelishvili, 1970):

$$B(t) = L_t \{Q_1(t), \dots, Q_m(t), W_1(t), \dots, W_1(t-\tau), \dots, \\ \dots, W_m(t), \dots, W_m(t-\tau)\} \quad (1.6)$$

where τ is the lag time between reservoirs. For a certain given realization over a time period T of the runoff process, the problem can be treated as a deterministic one, i.e. to find the minimum of:

$$\sum_{t=1}^T L_t \{Q_1(t), \dots, Q_m(t), W_1(t), \dots, W_1(t-\tau), \dots, W_m(t), \dots, W_m(t-\tau)\} \quad (1.7)$$

with respect to the control variables

$$W_i(t); i = 1, \dots, m; t = 1, \dots, T$$

On an ensemble of N realizations of the runoff process we shall find N such minima. If we make instead the optimization of the expected value, we shall find the minimum of the expression:

$$\sum_{l=1}^N \sum_{t=1}^T L_t \{Q_l^1(t), \dots, Q_l^m(t), W_l^1(t-\tau), W_l^m(t), \dots, W_l^m(t-\tau)\} \quad (1.8)$$

where l denotes the different realizations. The synthetic realizations of runoff are in the same way as before derived from its distribution functions.

The given examples of water management calculations assume that the full probability information of the natural runoff process is known. In fact hydrometric observations do not give such information. The present work, with its assumptions about the distribution functions of river runoff gives, however, the necessary characterization of this process for such calculations. This is true both in case when we use the analytical approach based directly on the distribution functions, as in the first examples, and when we use the Monte-Carlo method.

When this work was started in 1971 at the Geographical faculty of Moscow Lomonosov State University under the supervision of professor N A Kartvelishvili, it was ment to deal mainly with the problem of optimal regulation for hydropower production and risks involved. Just a smaller initial part was planned to describe stochastic models

for the runoff process. The application to Swedish conditions, where this kind of calculations has been used to a limited extent, made it necessary to perform a more thorough investigation of the distribution of river runoff and the way to simulate it. As the work was carried on, the interest has grown for this specific part and it was extended over whole Sweden. The intention was to see if any regional variations could be found in the fitted models. A study was also made on the procedure of estimation of parameters in the applied models. The material gathered, enabled the present publication.

2. NATURAL RIVER RUNOFF AS RANDOM PROCESS

The definitions in this chapter are mainly from Kartvelishvili (1967, 1970). Let us consider river runoff $Q = Q(t)$ at some point in a river as a function of time t . To characterize this process we use the probability distribution function of $Q(t)$ at the moment t :

$$P\{Q(t) < q\} = F_1(q, t) \quad (2.1)$$

which depends not only on q , but in the general case also on time t . It is apparent that the distribution function $F_1(q, t)$ can characterize the river runoff process completely only in the case, when the preceding events of the process do not influence the following ones.

Further information can be given by the bivariate distribution function i.e. the probability that at two following moments of time t_1 and t_2 , the conditions $Q(t_1) < q_1$ and $Q(t_2) < q_2$ will be satisfied:

$$P\{Q(t_1) < q_1, Q(t_2) < q_2\} = F(q_1, t_1; q_2, t_2)$$

Still more information will be given by the trivariate distribution function:

$$P\{Q(t_1) < q_1, Q(t_2) < q_2, Q(t_3) < q_3\} =$$

$$= F(q_1, t_1, q_2, t_2; q_3, t_3)$$

and so on. For the complete description of the river runoff process we need the infinite sequence of distributions $F_1, F_2, \dots, F_k, \dots$. It is obvious, that the series will be known if the distribution F_1 is given as well as the rule for the derivation of F_{i+1} from F_i .

When the distribution function F_k is given, all the functions for indices $i < k$ can be calculated from the following relation:

$$F_i(q_1, t_1; \dots; q_i, t_i) = F_k(q_1, t_1; \dots; q_i, t_i; \infty, t_{i+1}; \dots; \infty, t_k) \quad (2.2)$$

The right part of the formula above does not actually depend on $t_{i+1}, \dots, t_k$.

The river runoff process can be also represented by its conditional distribution function:

$$\begin{aligned} P \{Q(t_i) < q_i \mid Q(t_1) = q_1, \dots, Q(t_{i-1}) = q_{i-1}\} = \\ = f_i(q_i, t_i \mid q_1, t_1; \dots; q_{i-1}, t_{i-1}) \end{aligned} \quad (2.3)$$

The relation between the distribution function F_i and its corresponding conditional distribution function is described by:

$$\begin{aligned} F(q_i, t_i \mid q_1, t_1; \dots; q_{i-1}, t_{i-1}) = \\ = \frac{\int_{-\infty}^{q_i} f_i(q_1, t_1; \dots; q_{i-1}, t_{i-1}; z_i, t_i) dz_i}{f_{i-1}(q_1, t_1; \dots; q_{i-1}, t_{i-1})} \end{aligned} \quad (2.4)$$

where f_i is the probability density function corresponding to the distribution function F_i , i.e. $f_i = \partial^i F_i / \partial x_1 \dots \partial x_i$ and f_{i-1} in the same way is the density function corresponding to F_{i-1} .

For a random process with an infinite sequence of distribution functions, the function (2.3) will not possess for any value i , generally speaking, all the information about the process. Very important, however, are those cases for which the function F_i for some $i = n$ has this information. This means that the influence of events of the process earlier than t_1 does not exist, it exists only from events at moments $t_1, \dots, t_{i-1}$. In other words, for $i = n$ the following relation is satisfied:

$$\begin{aligned} P \{Q(t_n) < q_n \mid \dots; Q(t_{-1}) = q_{-1}; Q(t_0) = q_0; Q(t_1) = \\ = q_1; \dots; Q(t_{n-1}) = q_{n-1}\} \\ = P \{Q(t_n) < q_n; Q(t_1) = q_1; \dots; Q(t_{n-1}) = q_{n-1}\} \end{aligned} \quad (2.5)$$

In this case the process is called a Markov-process of the order $n-1$. For any $i > n$ we can now calculate F_i from F_n by the expression:

$$\begin{aligned}
 F_i(q_1, t_1; \dots; q_i, t_i) &= \\
 &= F_n(q_1, t_1; \dots; q_n, t_n) \cdot \\
 &\prod_{k=n+1}^i \frac{F_n(q_{k-n+1}, t_{k-n+1}; \dots; q_k, t_k)}{F_{n-1}(q_{k-n+1}, t_{k-n+1}; \dots; q_{k-1}, t_{k-1})} \quad (2.6)
 \end{aligned}$$

If the distribution function F_n is known for a Markov process all distribution functions F_i can be calculated, not only for $i \leq n$ (eq 2.2) but also for $i > n$ (eq 2.6).

Let $t_1 = t$, $t_2 = t + \tau_1$, $t_3 = t + \tau_2$, ..., $t_i = t + \tau_{i-1}$, where τ_k , $k=1, \dots, i-1$ can be even positive or negative numbers. Now we write down the distribution function F_i in the alternative way:

$$F_i(q_1, t_1; \dots; q_i, t_i) = F_i^X(q_1, \dots, q_i, t, \tau_1, \dots, \tau_{i-1}) \quad (2.7)$$

If all the functions F_i , $i = 1, 2, 3, \dots$ do not depend on t the considered process is a stationary one. Otherwise the process is nonstationary. Of importance for river runoff is the case of nonstationarity, when the distribution function F_i^X is a periodic function with the period τ , i.e.

$$\begin{aligned}
 F_i^X(q_1, \dots, q_i, t + \tau, \tau_1, \dots, \tau_{i-1}) &= \\
 &= F_i^X(q_1, \dots, q_i, t, \tau_1, \dots, \tau_{i-1}) \quad (2.8)
 \end{aligned}$$

This nonstationary process is called a harmonic process. The period of hydrologic processes is one astronomic year. We shall further postulate that the river runoff process can be considered to be ergodic. This means in brief that the average of the process can be calculated as the average over one realization in time. A more detailed definition will be the following. Let us consider some function Φ dependent on n values of $Q(t)$ at m different moments of

time $\Phi = \Phi \{Q(t_1), \dots, Q(t_m)\}$. We write down the following two means:

$$\Phi_s^{(j)} = \frac{1}{s} \sum_{k=0}^s \Phi \{Q^{(j)}(t_1 + kT), \dots, Q^{(j)}(t_m + kT)\} \quad (2.9)$$

$$\Phi_r = \frac{1}{r} \sum_{i=1}^r \Phi \{Q^{(i)}(t_1), \dots, Q^{(i)}(t_m)\} \quad (2.10)$$

where j and i are the order number of the realizations of the process. If there exists such T , that for any function Φ we have

$$\lim_{s \rightarrow \infty} \Phi_s^{(j)} = \lim_{r \rightarrow \infty} \Phi_r \quad (2.11)$$

independently on j , the process is harmonic with period T and ergodic. If (2.11) is valid for any T , the process is stationary and harmonic.

Up till now we have dealt with instantaneous runoff. In practice runoff is given as averages over some time period:

$$Q^x(t) = \frac{1}{\theta} \int_{t-\theta/2}^{t+\theta/2} Q(z) dz \quad (2.12)$$

In this work we shall use the period of one month. We could have chosen other time periods and expected good result. It would not, however, be advisable to use too small time periods. This also implies that river runoff from large catchments is used. The approach used here, to be further developed in the following, can not describe properly runoff from small catchments, where every outburst of rain results in a corresponding hydrograph. By a large catchment, thus should be understood a catchment, where the different storages within its area damps out individual rainfalls or shorter periods of snow-melt, and reflects instead the climatic changes at a larger scale.

We shall assume that the river runoff process for monthly average values can be approximated by a Markow process, which is harmonic, with the period one year, and ergodic. The process is thus described by twelve distribution functions:

$$F_n (q_t, t; q_{t-1}, t-1; \dots; q_{t-n}, t-n), t = 1, \dots, 12 \quad (2.13)$$

where n is the order of the Markov process.

We shall denote monthly average flow by Q_i , $i = 1, 2, \dots, 12$ excluding the asterisk in (2.12). For convenience we shall, however, in most cases use the module of runoff defined by $X_i = Q_i / \bar{Q}_i$, $i = 1, 2, \dots, 12$, where $\bar{Q}_i$ is the mean for the i -th month.

3. MARGINAL DISTRIBUTION FUNCTIONS FOR MONTHLY AVERAGE RUNOFF

3.1 Introduction

One dimensional distribution functions have been widely applied in hydrology for the analysis of stream flow observations. For a long time the interest was concentrated on frequency studies of annual runoff and annual floods. These phenomena are fairly well described by one dimensional distribution functions in most cases. The normal distribution function is fundamental and furthermore - a simple instrument for frequency studies, and was therefore very early applied to runoff analysis. It was soon shown, however, that the logarithm of flows gave a better fit to the normal distribution. Other laws of frequency distributions often used are the type I and type III of Karl Pearson's curves and the extreme value distribution curves. Still many others have been applied to frequency studies, quite a few of them, however, have been specially derived to suite hydrologic purposes. In some countries there have been intensive discussions about how to find the analytical form of a more generally applicable distribution, and the proper way to find it. It is not arbitrary which distribution one uses, as they give very different results when extrapolation is performed beyond the interval of observations (see for instance, Benson, 1968).

For the description of the probability characteristics for periods less than one year, say months and decades, it is necessary to abandon considering river runoff as a random variable and apply instead the theory of random processes. Onedimensional distribution functions frequently applied to describe the marginal distributions for months and decades are the Pearson type III distribution, the Kritskij and Menkel distribution and transformations of the normal distribution.

Theoretical determination of the distribution function for runoff phenomena is indeed a difficult task, as it involves complete utilization of the probability characteristics of the underlying processes. The a priori information of the form of the distribution curves is thus in most cases restricted to the facts that it should be non-negative and unbounded upwards.

It is obvious that for some period of time, limits could be set up for the magnitude of flow both upwards and downwards. Our knowledge about the nature of runoff tells us this, but we can not actually determine from this knowledge those limits. It seems to be more reasonable to use distributions that give small probability of occurrence at hypothetical bounds, than to determine really the boundary by statistical methods with the uncertainty they involve. This must be specially significant if extreme events are of importance. If the average behaviour is of more interest as for instance in regulation calculations, bounded distributions could be accepted.

A more difficult question is whether the distribution function is unimodal or not. When considering the distribution for monthly and decade runoff for some periods of the year, the possibilities that the distribution curves can have a more complicated form could not be excluded. One such period is spring when in some years the flow is mostly fed from groundwater storages, while in other years there is also runoff from snowmelt. Unimodality is assumed because of the lack of knowledge.

The genetical information thus gives us only vague ideas of the analytical form of the distribution functions. The final choice should be based on observations over the random variable, i.e. in our case-observations of monthly average flow. This procedure has two stages. The first one is to determine the analytical form of the distribution and the second one - to evaluate the parameters of this analytical distribution. Statistical methods enables us then to make a test, based on the observations, whether the hypothesis, that the analytical distribution with its calculated parameters fits the observations, could be rejected or not. Such test does not give, of course, any theoretical justification for the chosen analytical form, but only a reliable empirical formula. This often gives the result that two or more analytical distributions can be found to be equally good, both from genetical and statistical reasoning. The final choice will be based then on practical reasons, such as the amount of computing work for determination of parameters. A way to avoid this problem, is to find regional justification for the analytical distribution. The analytical distribution should not be tested thus only at one site in a region, but at several sites.

We get guidance for the choice of the analytical distribution function from the empirical functions of the sample and its moments. It is obvious that for a territory as large as Sweden we shall observe large differences in these due to differences in the physiographic conditions. This implies that only one analytical distribution function is, may be, not flexible enough to consider all situations. Certain analytical distributions can be better for certain physiographic conditions. It is hardly worth striving for to find one universal form. We know that it can be easily found if we only include enough parameters for the description of river runoff. The accuracy in their determination is then reduced. Then, it must be better study the physiographic conditions, as well as regularities in the empirical distributions, and let this guide us to an analytical distribution with maybe only regional application.

In the following part of this chapter we shall first look at empirical probability distribution functions and try to describe their regional variations. With this as a base, we shall discuss applicable analytical probability distribution functions. More thoroughly will be treated distribution functions, based on transformations of the normal distribution. Formulas for the calculation of the parameters of the discussed analytical distributions will be derived with different methods of estimation. Finally, conclusions about applicable analytical distribution functions for monthly average runoff will be given.

3.2 Empirical distributions

Let us consider the population of all outcomes of the module of average river runoff X for some certain month. This is described by its one dimensional probability distribution function $F_1(x)$. It should be noted, that in hydrologic application the analytical form of the distribution function is actually not known, as was mentioned above. Given is N observations of the module values from this population, ordered from least to greatest:

$$x^1 < x^2 < \dots < x^i < \dots < x^N$$

The relative frequency $F_i^X(x) = i/N$, where i is the number of observations less or equal x_i , approximate the probability $p = P(X \leq x^i) = F(x^i)$ by the law of large numbers. It is called the sample distribution function or as we shall use here - the empirical distribution function. $F^X(x)$ is a step function with jumps at the points $x = x^i$, $i = 1, \dots, N$. Further, we have that $F^X(x) = 0$ for $x < x^1$ and $F^X(x) = 1$ for $x > x^N$. For small samples, which is usually the case in hydrologic practice, $F^X(x)$ plotted against x can be only expected to give rough ideas about the underlying distribution function $F(x^i)$ (cf. 3.2.1).

3.2.1 Empirical moments

The empirical distribution function is defined by giving the weight $1/N$ to each observation. For its moments we have thus the following expressions:

$$a_k = \frac{1}{N} \sum_{i=1}^N (x^i)^k = \sum_{-\infty}^{\infty} x^k dF^X(x) \quad (3.1)$$

$$m_k = \frac{1}{N} \sum_{i=1}^N (x^i - a_1)^k \quad (3.2)$$

which are the empirical moment of the origin of order k and the central moment of order k , respectively. These are approximations to the moments α_k and μ_k of the population distribution $F(x)$. The empirical moments are subject to random errors, which are the larger the smaller N is. Furthermore, they are systematically biased. This can be caused by the fact that we have a finite sample, that the x 's are correlated and that the observations are skewed. When N is small, as in hydrologic practice, these facts are of great importance. They will be dealt with below, when discussing parameter estimation (section 3.5).

The low order empirical moments can characterize to a large extent the empirical distribution function. They are therefore useful instruments for the conclusions about the regional differences in the run-off distributions. Due to the small samples, it is first of all advisable not to use higher order of the moments than two, or may be three, for such studies. The error in the third order moment can, how-

ever, be large. One single very large observed value can, for instance, give predominant influence on this. We should thus be very cautious with the conclusions we draw from this parameter.

We shall not use here directly the second and third order central empirical moments, but dimensionless forms of these, the variation coefficient C_V and the skewness coefficient C_S defined by:

$$C_V = \sqrt{m_2}/a_1 \quad (3.3)$$

$$C_S = m_3 / \sqrt{m_2^3} \quad (3.4)$$

3.2.2 Order statistics

The quantity $p = P(X \leq x) = F(x)$ can be said to be drawn from a rectangular distribution. The probability that we shall observe a value less than the i -th of the ordered observations

$$P(X < x^i) = p_i$$

is called the i -th order statistic. On conditions of independancy the probability density function for p_i is

$$f_{p_i}(p) = \frac{N!}{(i-1)!(N-i)!} p^{i-1} (1-p)^{N-i}$$

This is the density function for the beta-distribution. The expected value of p_i is:

$$E(p_i) = i / (N+1) \quad (3.5)$$

and for the variance of p_i the following expression is given:

$$\text{Var}(p_i) = i(N-i+1) / \{(N+1)^2(N+2)\} \quad (3.6)$$

The expected value of p_i can be used for graphical representation of the empirical distribution $F^X(x)$, when it is plotted against corresponding values of x^i . For $x = x^1$ and $x = x^N$ we have then that $F^X(x^1) = E(p_1) = 1 / (N+1)$ and $F^X(x^N) = E(p_N) = N / (N+1)$, respec-

tively, $F^X(x)$ for a finite sample does not obtain the value neither for the certain event nor for the impossible event. This is not the case when we use the relative frequency. Eq (3.5) should be thus formally a better expression to use. Further necessary properties for plotting positions are given in Gumbel (1958) and will not be discussed here.

- Let us now consider the inverse problem when a set of p_i s from a rectangular distribution is given. The order statistic:

$$x^i = F^{-1}(p_i),$$

where F^{-1} is the inverse to F , is called a transformed β -variable (Blom, 1958). It has the density function:

$$f_{x^i}(x) = \frac{N!}{(i-1)!(N-i)!} \{F(x)\}^{i-1} \{1-F(x)\}^{N-i} f(x),$$

where $f(x) = dF(x)/dx$. It can be shown that $F_{x^i}(x^i) = F_{p_i}(p_i)$. From this follows that p_i and x^i are corresponding points (Weibull, 1961). The expected value u^i of x^i is given by:

$$u^i = E(x^i) = \int_{-\infty}^{\infty} x f_x(x) dx = \int_0^1 F^{-1}(p) f_p(p) dp$$

We note that the expected value is dependent on the population distribution $F(x)$. An approximate formula is:

$$u^i = E(x^i) = F^{-1}(p_i) \quad (3.7)$$

where p_i is dependent on i , N and the distribution $F(x)$. Blom (1958) has suggested the general expression:

$$p_i = \frac{i - a'}{N - a' - b' + 1}$$

where the coefficients a' and b' will be different for different distributions. For the normal distribution Blom gives the following formula:

$$p'_i = (i - 3/8) / (N + 1/4) \quad (3.8)$$

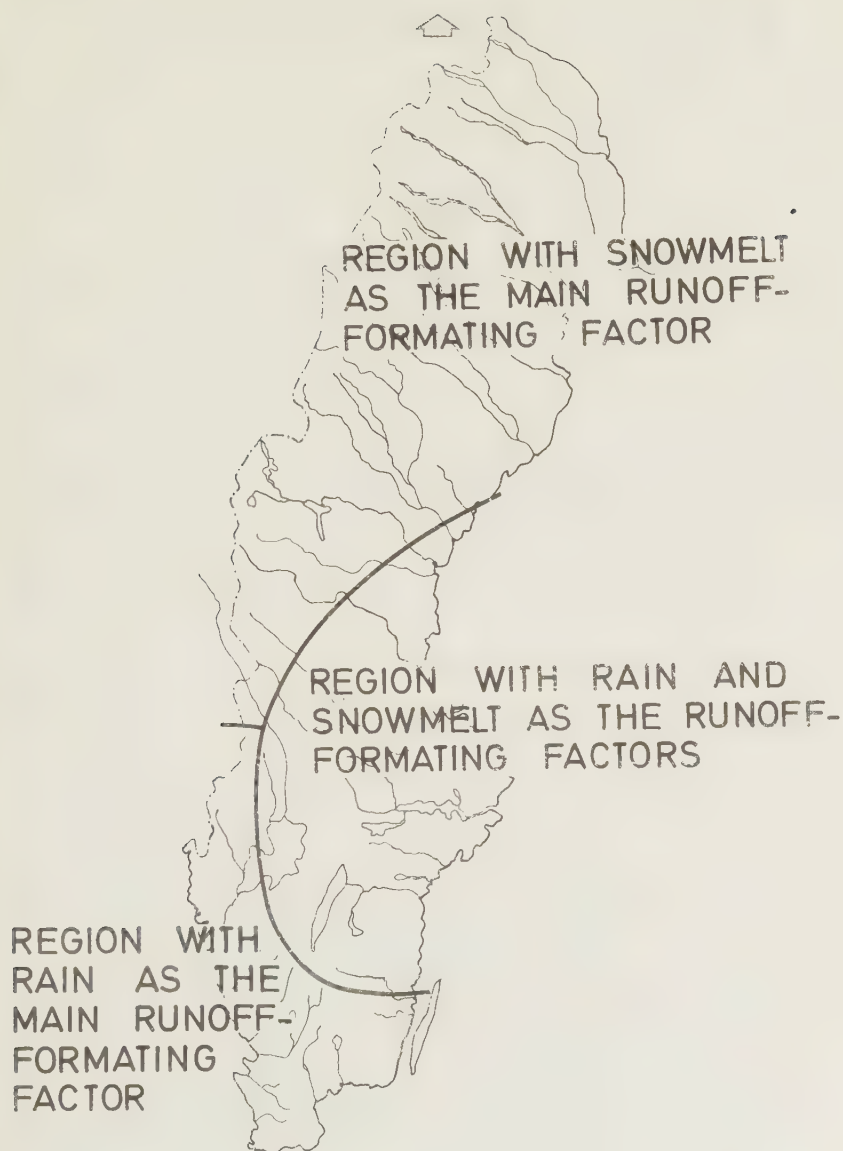


Fig 3.1 Map over Swedish regions with equal runoff forming factors (after Kuprijanov, 1960).

Exact values of p_i for the normal distribution can be found in Pearson (1931 - 32). Alekseev (1963) for the same purpose has used the formula:

$$p''_i = \frac{i - a''}{N + b''} \quad (3.9)$$

where

$$a'' = (N + 1 - (N(N+1))^{1/2}) / 2 \approx 0.25$$

$$b'' = (N(N+1))^{1/2} - N \approx 0.5$$

It is derived so that p_i preserves the expected value and the variance of the rectangular distribution. Values of $u^i = \Phi^{-1}(p''_i)$, $i = 1, \dots, N$, where Φ^{-1} denotes the inverse to the normal distribution function (cf. eq. (3.12)), have been tabulated for different N by Alekseev (1971). The corresponding type of table was prepared for the present work, but instead of p''_i , p'_i has been used. We shall use this table below to derive empirical transformation curves for the normal distribution.

3.3 Regional variations in empirical distribution of the module of average monthly runoff

Regional differences in the empirical distribution of monthly module values are explained by the variations of the physiographic conditions, i.e. climate, topography, geology, pedology, biogeography and land use. Schematic division of Sweden into regions of the same runoff regimes which includes some enumerated factors, though not all, has been done by Kuprijanov (1960) Fig 3.1 and by Melin (1970) Fig 3.2. The two maps were based on the climatic runoff forming factors: snow- and icemelt and precipitation. Melin makes division into two main regimes: a nival mountain and lowland regime for the north of Sweden and an atlantic mixed snow and rain regime in the south. Kuprijanov has made a further division of this last region into two. One in the east, where the runoff is mainly fed by snowmelt, and another one in the western part of this region, where precipitation is the main factor. Melin (1970) has also made a more detailed division of

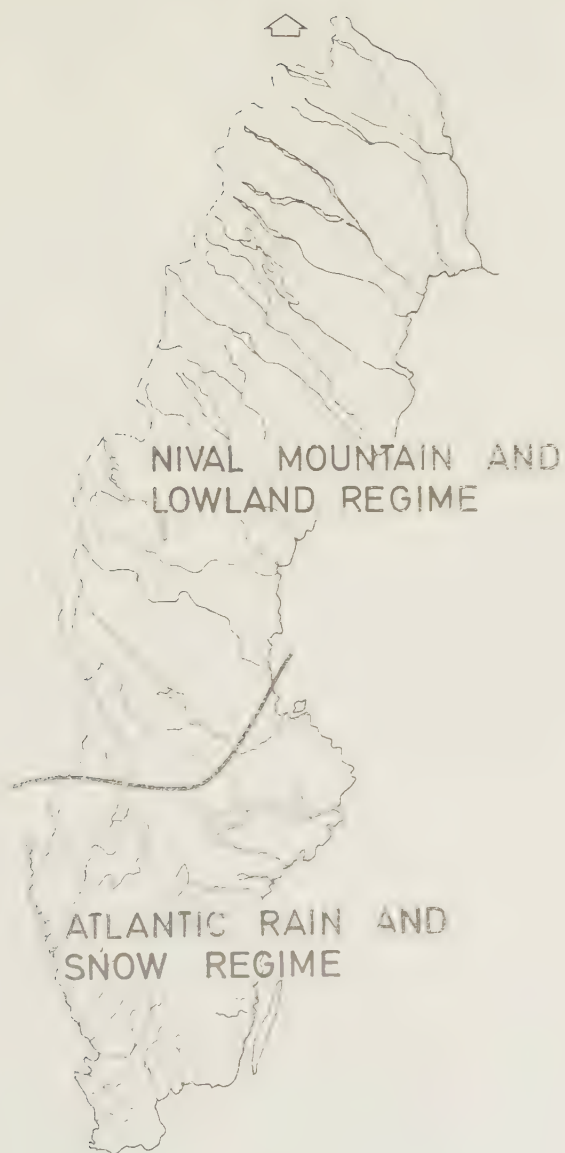


Fig 3.2 Map over Swedish river regimes (after Melin, 1970).

Sweden into hydrologic regions, where he also considers regularities in trends of 30-year observation series of annual runoff, the magnitude of the mean annual runoff and the geographical location (Fig 3.3). The general subdivision of Swedish rivers by Melin is the following:

- A. Polar Rivers
- B. Northern Mountain Rivers
- C. Central Mountain Rivers
- D. Mountain and Forest Rivers in the south of
Norrrland and north of Svealand
- E. Forest and Coastal Rivers in Norrrland
- F. The Rivers of Southern Sweden
 - a. eastern rivers
 - b. western rivers
- G. Rivers from the large Swedish lakes
- H. Rivers with mixed regime

To exemplify the large differences in the formation of river runoff from various parts of Sweden monthly runoff observation series from nine different rivers are shown (Fig 3.4 - 6). The location of each observation site can be found on the map Fig 3.7. All three series in Fig 3.4 are from the north of Sweden i.e. the main runoff forming factor is snowmelt. The series 28-444 Hällbacken in the river Vindelälven are the most extreme ones, where snowmelt is totally dominant. The observation site is situated in the mountain region. Coming down to the lowland, the spring peak is not that extreme and further, in some years a small peak is observed due to rainfall in autumn (1-589 Kallio). This is still more pronounced with movement southwards (28-1302 Degerfors). The two series 108-274 Edebäck and 53-121 Fäggeby in Fig 3.5 are also situated in the region, where snowmelt is the main runoff forming factor, but close to the region of mixed regimes. The peaks due to snowmelt and rainfall, respectively can be of the same magnitude for some years. The series 67-154 Motala are very specific as this river comes from the outflow of lake Vättern. No extreme peaks are found and even the annual period is hard to distinguish. In Fig 3.6 all the series are from the South of Sweden. The peaks can be observed at any time, except for summer. Compared to the series from the North they seem to be very irregular.

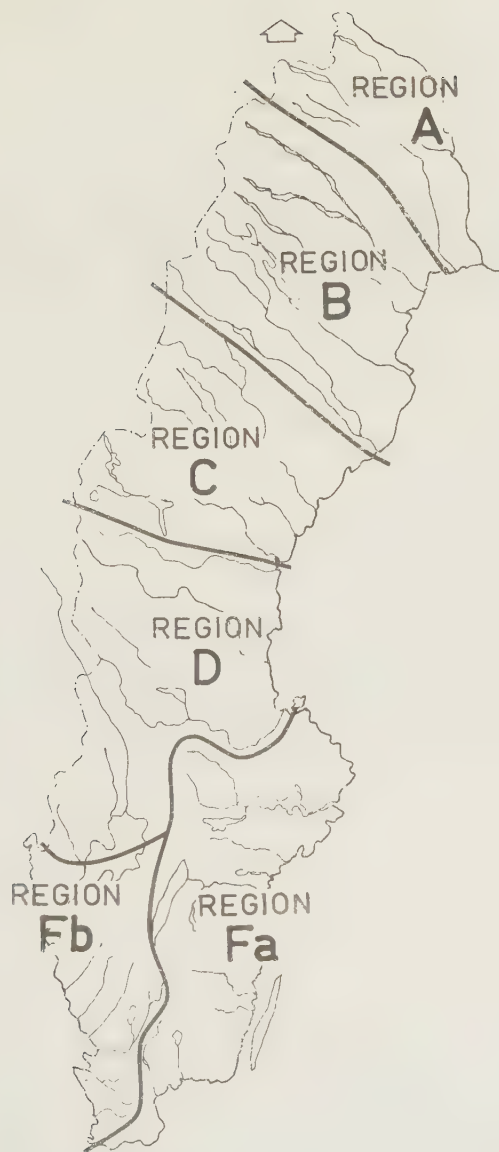


Fig 3.3 Map over Swedish hydrological regions (after Melin, 1970).

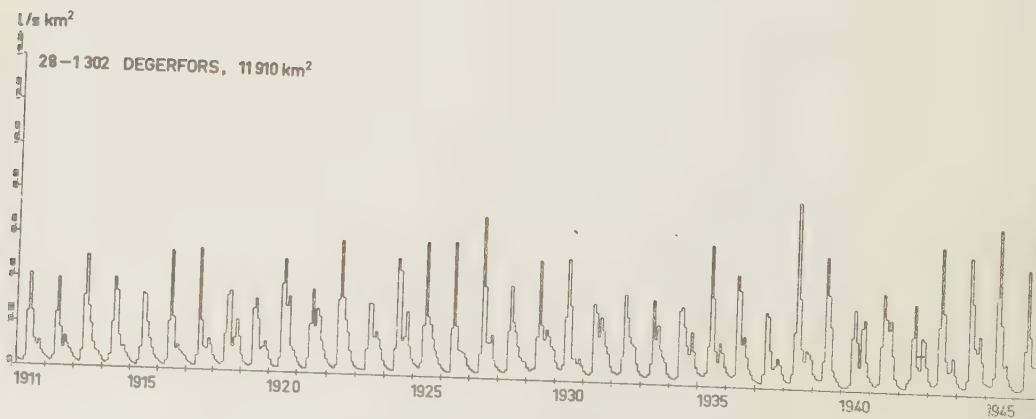
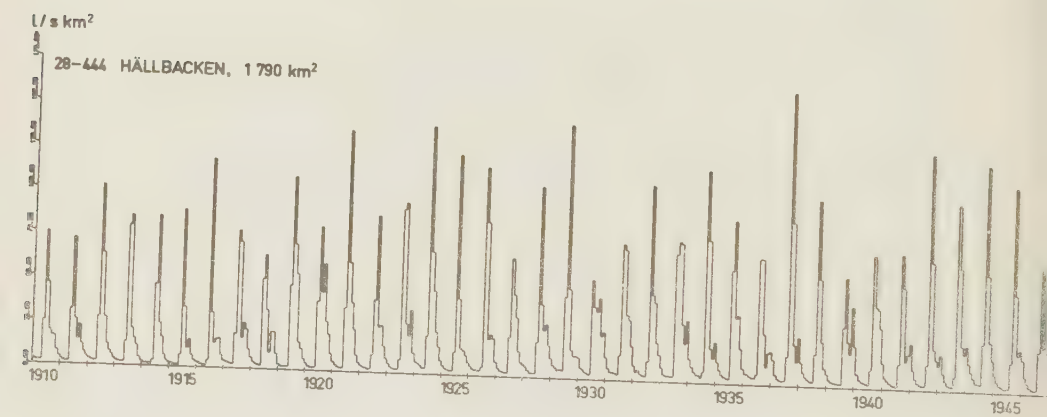
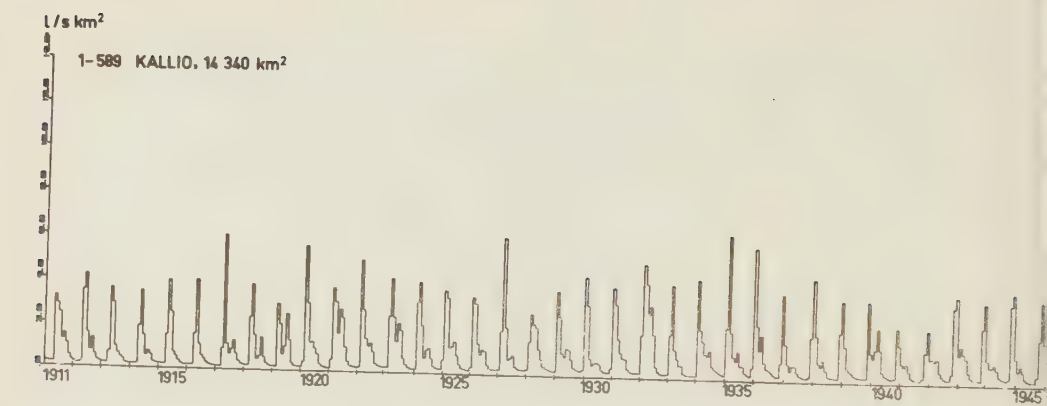


Fig 3.4 Observed runoff series from the northern Sweden.

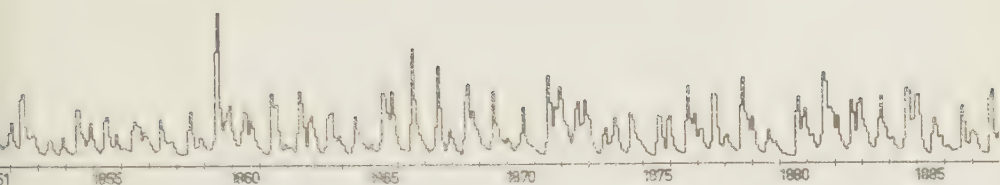
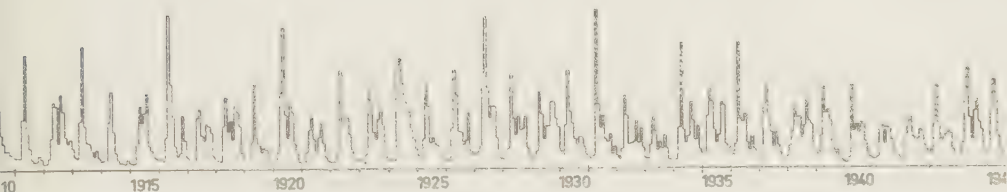
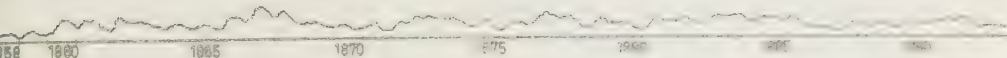
5 km²3-121 FÄGGEBY, 25 300 km²km²100-274 EDEBÄCK, 8 560 km²km²67-154 MOTALA, 6 360 km²

Fig 3.5 Observed runoff series from the central Sweden.

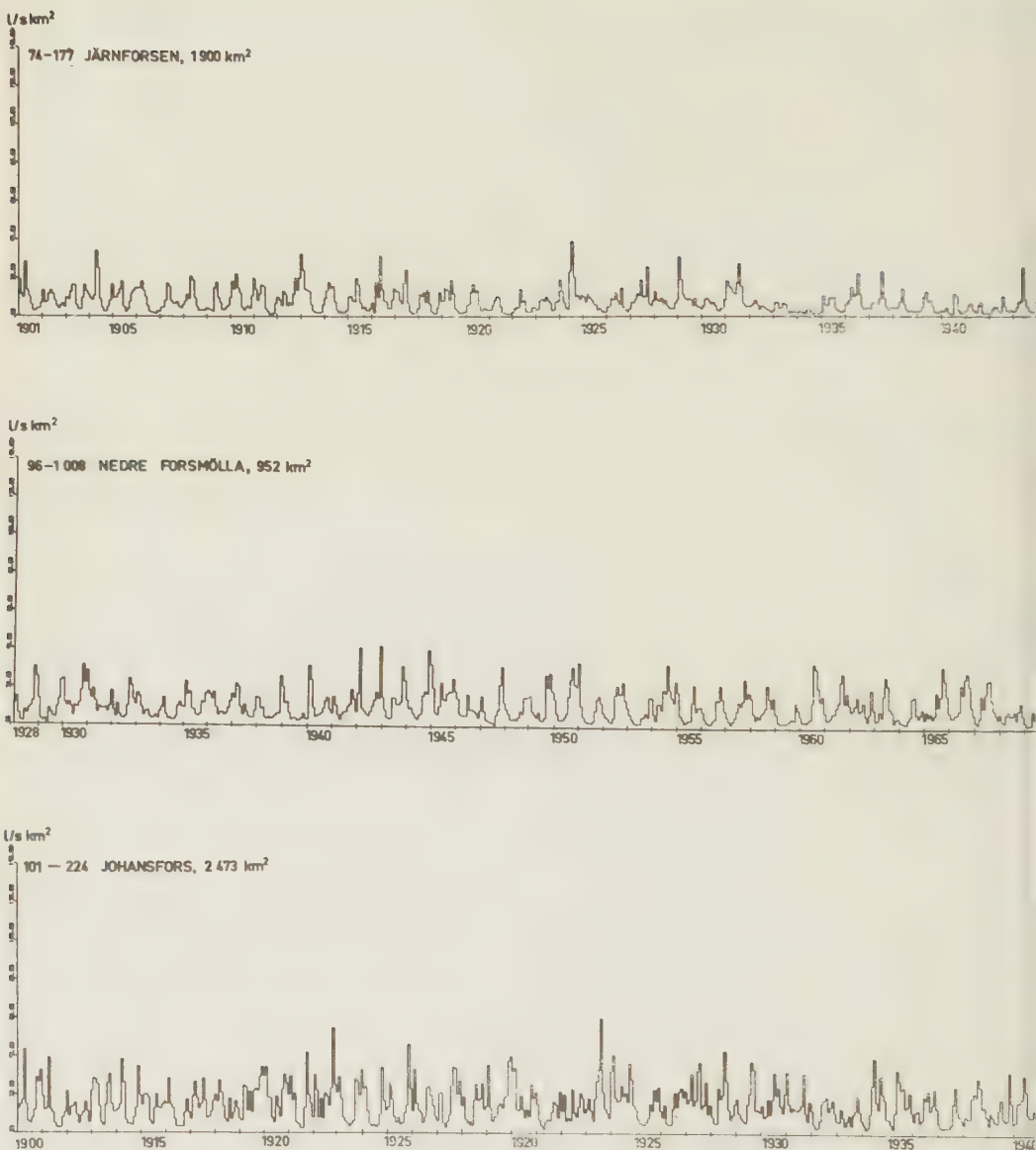


Fig 3.6 Observed runoff series from the southern Sweden.

Now we shall study the regional variations in the empirical runoff distribution in more detail. For this we shall accept the latter subdivision. It does not answer precisely to our understanding. For instance, we should have suggested to divide northern regions into mountain and lowland regions. Further, more physiographic factors should be included, which will give a more detailed map. We shall not, however, develop this topic any more, as it represents a theme for a separate study.

For the discussion of regional characteristics of the runoff distribution functions it is reasonable to base this discussion on the moments of these distributions. Examples of the variations in these moments will be shown below to give a general pattern. For a final investigation of these parameters the number of observation series should be, of course, larger.

3.3.1 Regional variations in empirical moments of monthly average runoff

Mean values, variation coefficients and skew-coefficients have been calculated for 38 series of monthly runoff. The locations of the observation sites are shown in Fig 3.7. The series are taken from Melin (1955). Most of the series have been completed with data from the Swedish Meteorological and Hydrological Institute. Name of observation point, name of the river and period of observations are given in table 3.1.

3.3.1.1 Polar rivers

Calculated parameters for the rivers in this group are given in tables 3.2 - 4 and two examples are shown in Fig 3.8 and 3.9. It is advantageous to distinguish observation points in the mountain region and those downstreams in the lowland area. The main runoff forming factor is snowmelt, and the maximum mean discharge values are therefore observed during the snowmelt period June and July for the mountains and May and June for the lowland. During summer and autumn we have the recession of the flood, with some increase in the mean values

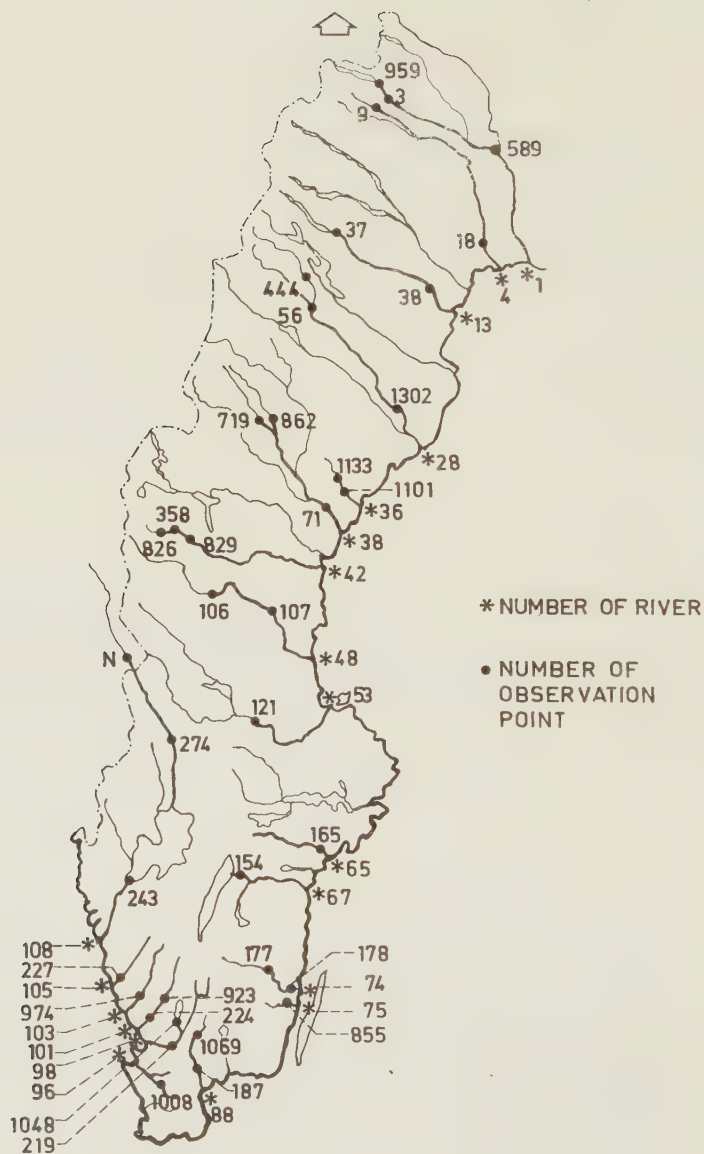


Fig 3.7 Map over used observation points.

TABLE 3.1

Observation data used					
Number of the station	Name of the observation station	Name of the main river	Period of observations	Length of observations	Catchment area sq km
1-3	Jukkasjärvi	Torneälv	1915 - 67	53	5999
1-589	Kallio	Torneälv	1911 - 71	61	14340
1-959	N Abiskojokk	Torneälv	1915 - 72	58	3294
4-10	Fjällåsen	Kalixälv	1917 - 68	51	2260
4-18	Mojärv	Kalixälv	1917 - 58	41	17712
13-37	Stenudden	Piteälv	1916 - 41	26	2440
13-38	Ålvsbyn	Piteälv	1910 - 41	32	10593
28-56	Sorsele	Umeälv	1910 - 71	62	6110
28-444	Hällbacken	Umeälv	1910 - 71	61	1790
28-1302	Degerfors	Umeälv	1911 - 49	39	11910
36-1101	Mellånsel	Moälven	1922 - 50	28	1452
36-1133	Kubbe	Moälven	1923 - 50	27	775
38-71	Forsmo	Angermanälven	1909 - 38	31	21490
38-719	Tåsjön	Angermanälven	1913 - 45	32	2599
38-862	Rörström	Angermanälven	1913 - 51	38	2469
42-358	Ø Torsborg	Ljungan	1916 - 67	51	1364
42-826	Storsjö	Ljungan	1915 - 63	47	928
42-829	Fotingen	Ljungan	1915 - 69	54	2592
48-106	Sveg	Ljusnan	1914 - 67	54	8490
48-107	Ljusdal	Ljusnan	1909 - 67	59	13720
53-121	Fäggeby	Dalälven	1831 - 1919	68	25300

(Continued)

Number of the station	Name of the obser- vation station	Name of the Main river	Period of observations	Length of observations	Catchment area sq km
65-148	N Täckhammare	Nyköpingsån	1909 - 38	24	3578
67-154	Motala	Motala ström	1858 - 1919	62	6360
74-177	Järnforsen	Emån	1901 - 60	60	1900
74-178	Klämma	Emån	1909 - 35	27	4170
75-855	Getebro	Alsterån	1920 - 71	52	1340
88-187	Hönjebro	Helge å	1909 - 50	42	2160
88-1069	Möckeln	Helge å	1921 - 71	50	1010
96-1008	N Forsmöllan	Rönne å	1928 - 71	44	952
98-219	Ashult	Lagan	1909 - 50	42	5480
98-1048	Bolmen	Lagan	1909 - 38	30	1650
101-224	Johansfors	Nissan	1900 - 43	44	2473
101-923	Färgebro	Nissan	1918 - 41	24	1650
103-974	Kila	Åtran	1911 - 38	28	2518
105-227	Asbro	Viskan	1909 - 50	42	2160
108-	Nybergsund	Klarälven	1909 - 50	42	4420
108-243	Sjötorp	Göta älv	1807 - 1970	164	46830
108-274	Edebäck	Klarälven	1910 - 47	38	8560

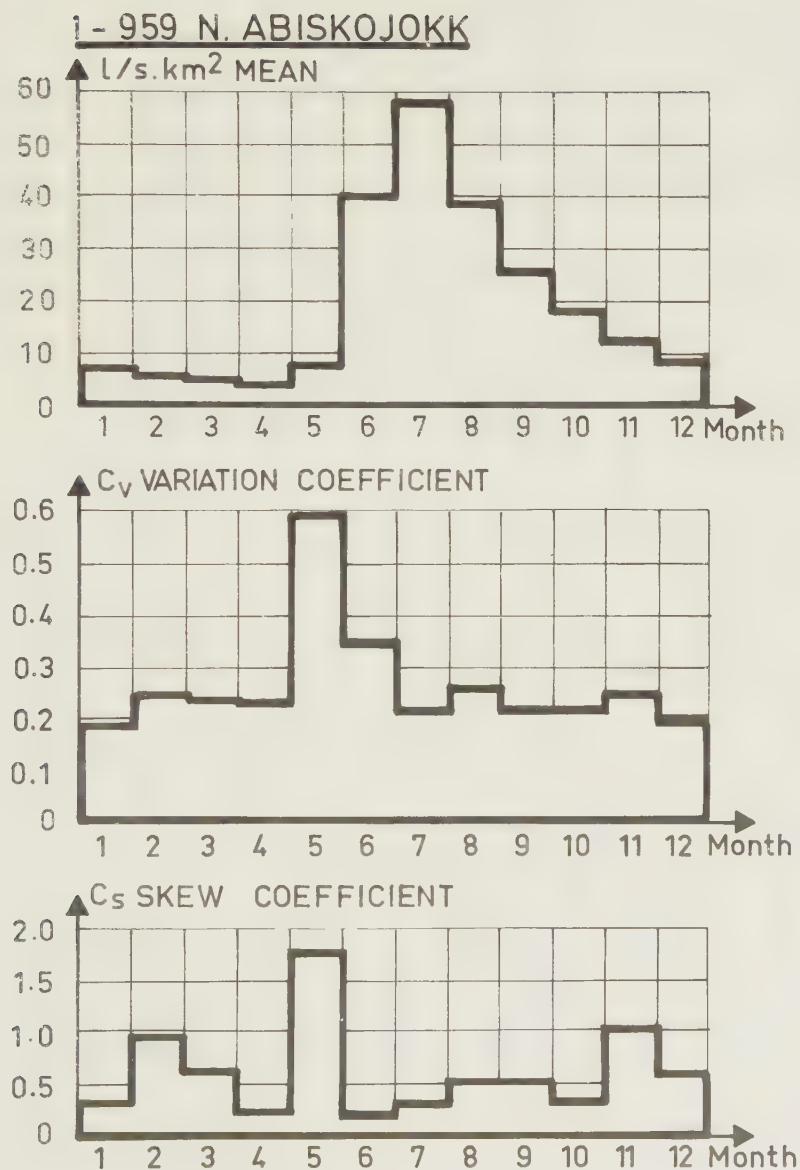


Fig 3.8 Graphs of monthly means and coefficients of variation and skewness for region A (mountain).

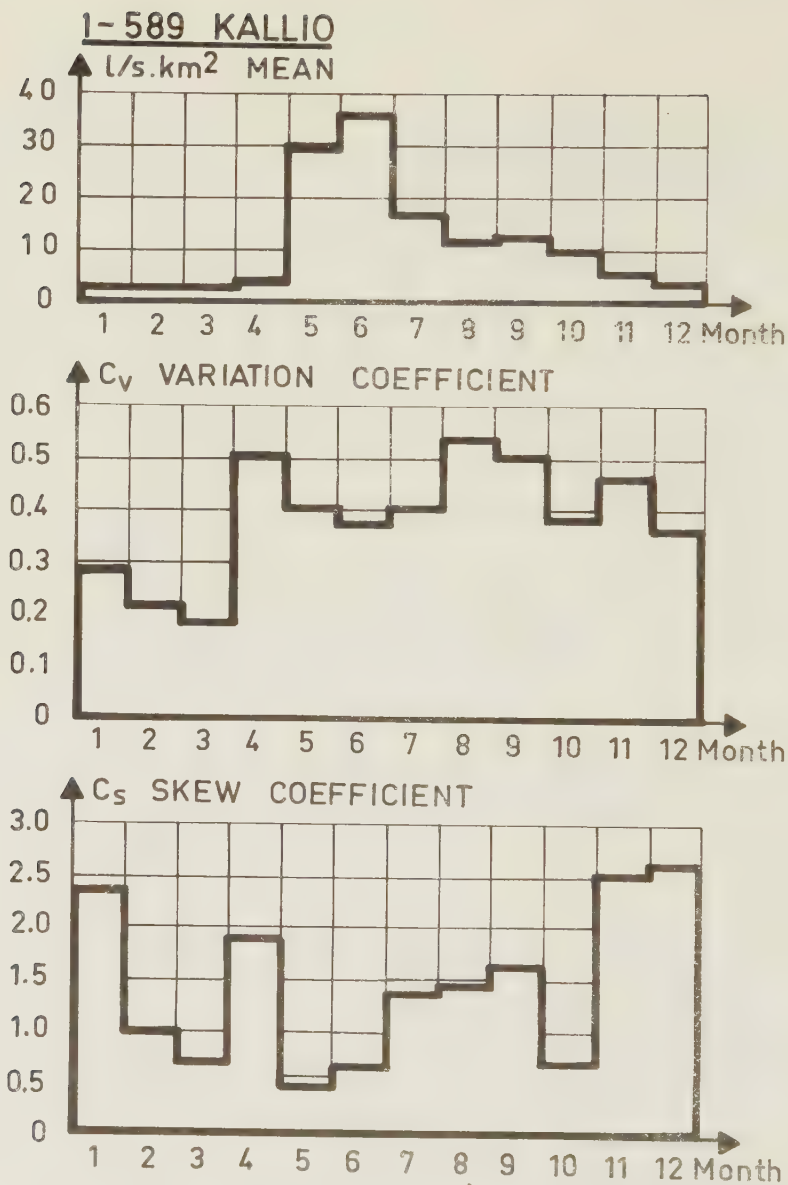


Fig 3.9 Graphs of monthly means and coefficients of variation and skewness for region A (lowland).

during late summer and autumn in the downstream area. Winterflow is very low. The climatic conditions in respect to runoff are stable from year to year, with the exclusion for the snowmelt period in the mountain area. This is indicated by the relatively small variation coefficient. In the lowland we observe large values of the variation coefficient, not only during snowmelt, but also in late summer and autumn. The skew coefficient does not show any regular pattern throughout the year. This can probably be caused by large statistical errors in this parameter. There is, however, a tendency that large values in the variation coefficient give corresponding large values in the skew coefficient, but it is not very well pronounced.

According to the variations in the mean and the variation coefficient, we can distinguish the following periods for mountain observation points:

- A 1: Winter flow; low mean values and low variation coefficient.
- A 2: Beginning of snowmelt; relative low mean values and high coefficient of variation.
- A 3: Snowmelt peak; very high mean values and variation coefficient moderate or relatively low.
- A 4: Recession; mean values moderate and variation coefficient low.

In the lowland we have to add a fifth period during late summer and autumn. It has moderate mean values and high variation coefficient (A 5).

TABLE 3.2 Monthly mean runoff values for region A

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1-3	5.1	4.1	3.5	3.1	12.3	40.3	46.3	32.5	22.5	15.8	9.8	6.8
1-589	2.8	2.5	2.2	3.1	29.5	35.2	16.1	11.5	11.4	9.4	5.3	3.6
1-959	6.3	5.4	4.5	4.2	7.5	40.0	58.2	38.5	25.5	18.2	12.1	8.1
4-10	3.9	3.0	2.6	3.0	31.4	60.1	46.0	32.3	22.5	14.1	8.4	5.7
4-18	4.2	3.5	3.2	7.3	37.7	40.5	29.8	23.0	19.5	16.4	18.7	5.5

TABLE 3.3 Monthly variation coefficients for region A

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1-3	0.18	0.19	0.20	0.19	0.56	0.28	0.20	0.24	0.21	0.18	0.23	0.21
1-589	0.28	0.21	0.18	0.50	0.40	0.37	0.40	0.53	0.50	0.38	0.46	0.36
1-959	0.18	0.24	0.23	0.23	0.59	0.34	0.21	0.25	0.21	0.21	0.24	0.19
4-10	0.23	0.22	0.22	0.33	0.49	0.33	0.24	0.36	0.29	0.32	0.28	0.33
4-18	0.24	0.19	0.18	0.53	0.40	0.36	0.23	0.44	0.38	0.33	0.45	0.40

TABLE 3.4 Monthly skew coefficients for region A

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1-3	0.52	0.90	0.77	0.50	0.52	0.39	0.56	0.68	0.61	0.07	1.21	1.35
1-589	2.29	0.95	0.63	1.83	0.44	0.61	1.33	1.40	1.57	0.65	2.48	2.54
1-959	0.30	0.93	0.57	0.22	1.71	0.16	0.29	0.50	0.51	0.31	1.01	0.58
4-10	0.56	0.27	0.13	0.57	0.39	0.26	0.02	0.78	0.30	0.58	0.57	2.60
4-18	1.39	0.58	0.16	2.58	0.68	1.25	0.55	2.07	0.82	0.32	0.95	2.68

3.3.1.2 Northern Mountain Rivers

Calculated coefficients for this group of rivers are given in tables 3.5 - 7 and one example is shown in Fig 3.10. The general behaviour of the variation in the three calculated moments throughout the year is the same as for the previous region. The peak of the mean value during snowmelt is still more expressed. This is also valid for the variation coefficient in the first month of snowmelt. A general increase in the variation coefficient is noticeable. The tendency of higher values of this parameter during late summer and autumn occurs not only in the lowland but also in the mountains. The skew coefficient has less very high values. The same five periods as for the previous region can be distinguished (B 1 - B 5)

TABLE 3.5 Monthly mean runoff values for region B

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
13-37	5.3	4.5	3.6	3.2	17.6	84.8	80.3	44.6	28.2	20.9	12.7	7.7
13-38	5.1	4.1	3.5	5.0	26.1	37.8	30.9	20.7	16.8	14.2	10.2	7.0
28-56	4.0	3.2	2.7	2.7	28.8	80.1	41.7	24.8	19.9	17.0	9.9	5.7
28-444	4.4	3.3	2.7	2.7	24.5	94.9	53.0	26.8	20.1	16.7	9.4	5.5
28-1302	4.2	3.3	2.9	4.6	23.1	49.0	29.4	17.5	14.7	14.6	10.3	6.4

TABLE 3.6 Monthly variation coefficients for region B

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
13-37	0.25	0.28	0.32	0.37	0.77	0.34	0.28	0.36	0.30	0.32	0.32	0.22
13-38	0.36	0.38	0.26	0.40	0.45	0.32	0.22	0.36	0.38	0.37	0.34	0.30
28-56	0.23	0.22	0.20	0.26	0.61	0.34	0.30	0.47	0.43	0.42	0.37	0.2
28-444	0.34	0.30	0.26	0.29	0.72	0.32	0.38	0.47	0.35	0.39	0.36	0.37
28-1302	0.33	0.31	0.33	0.38	0.47	0.30	0.27	0.38	0.39	0.43	0.41	0.40

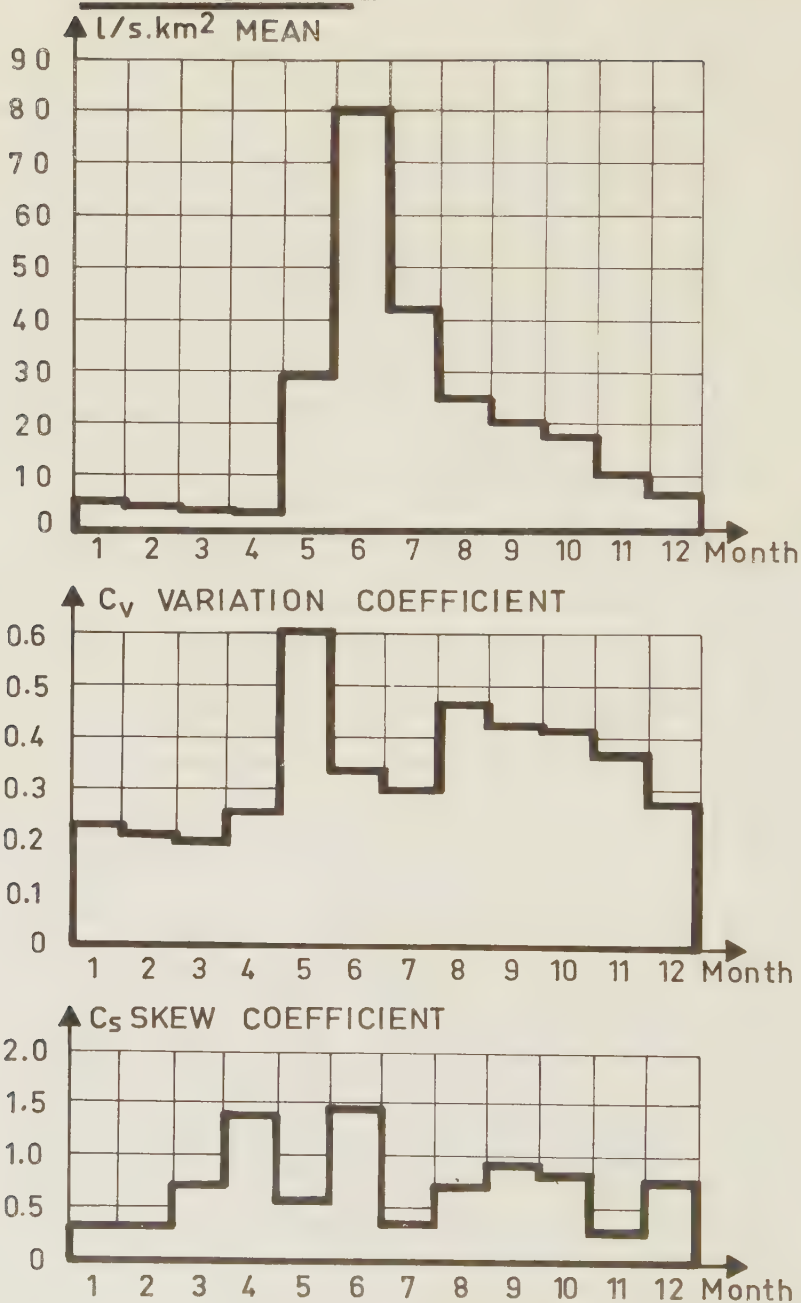
28-56 SORSELE

Fig 3.10 Graphs of monthly means and coefficients of variation and skewness for region B.

TABLE 3.7 Monthly skew coefficients for region B

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
13-37	0.33	0.77	0.69	0.87	0.91	0.52	0.43	0.14	0.87	0.80	0.96	0.80
13-38	2.96	3.50	0.95	1.30	1.28	1.10	0.34	0.69	1.12	0.62	0.42	0.53
28-56	0.29	0.30	0.72	1.37	0.54	1.44	0.33	0.69	0.90	0.82	0.28	0.72
28-444	1.25	0.91	0.70	1.26	1.07	0.22	0.46	0.77	1.04	0.79	0.72	1.48
28-1302	0.88	0.60	0.71	0.86	0.42	0.67	0.34	0.59	1.32	1.02	0.88	1.34

3.3.1.3 Central Mountain Rivers

The coefficients are given in tables 3.8 - 10 and one example is shown in Fig 3.11. The general pattern for the average value is still valid. The second peak during autumn is still more marked than in the two previous regions. This has also effect upon the variation coefficient, which is as high in autumn as at the beginning of snowmelt. Now this coefficient has minimum during winter and summer. No clear difference can be seen between the mountain area and the lowland. The five periods (C 1 - C 5) are not that distinct as for the previous regions.

38 - 71 FORSMO

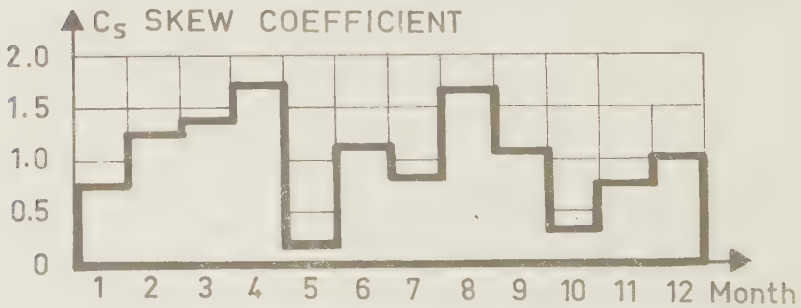
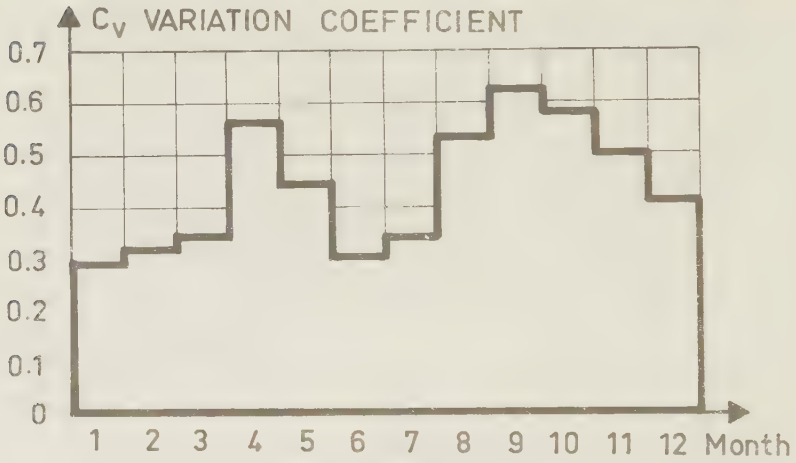
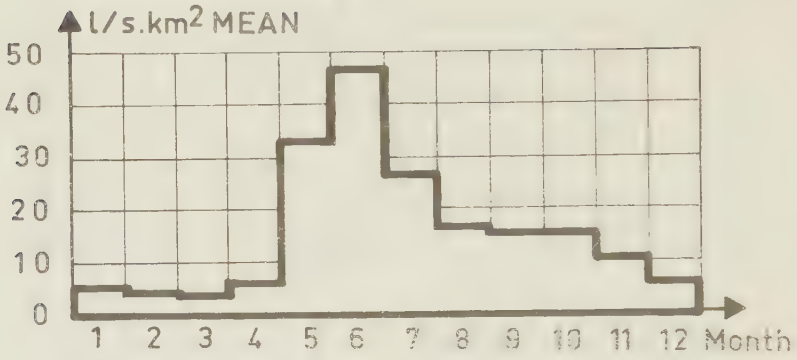


Fig 3.11 Graphs of monthly means and coefficients of variation and skewness for region C.

TABLE 3.8 Monthly mean runoff values for region C

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
38-71	4.5	3.7	3.3	5.3	32.4	46.9	26.1	16.1	15.1	15.2	10.7	6.5
38-719	6.2	5.4	4.6	5.8	46.2	72.4	33.8	20.0	20.4	23.1	13.8	8.5
38-862	4.8	4.1	3.6	5.7	40.9	40.9	18.6	13.3	13.8	15.0	10.9	7.3

TABLE 3.9 Monthly variation coefficients for region C

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
38-71	0.29	0.31	0.34	0.56	0.44	0.30	0.34	0.53	0.62	0.58	0.50	0.41
38-719	0.34	0.48	0.47	0.58	0.46	0.27	0.37	0.54	0.50	0.48	0.40	0.35
38-862	0.34	0.36	0.45	0.65	0.45	0.38	0.35	0.63	0.55	0.52	0.39	0.36

TABLE 3.10 Monthly skew coefficients for region C

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
38-71	0.71	1.42	1.34	1.68	0.15	1.09	0.79	1.62	1.05	0.28	0.72	1.00
38-719	0.57	2.00	1.35	1.71	0.08	0.23	0.45	1.99	0.57	0.20	0.03	0.35
38-862	0.66	0.76	1.22	2.01	0.22	1.02	0.84	3.11	0.78	0.32	0.39	0.53

3.3.1.4 Mountain and Forest Rivers in the South of Norrland and North of Svealand

Mean, variation and skew coefficients are given in Table 3.11 - 13 and one example of the variations in these coefficients is shown in Fig 3.12. No larger differences can be seen conferred to the previous region. The variation coefficient is very high especially in the western part of the region (108-N, 108-274) and not only during autumn but also in early winter. It does not change, however, the subdivision into periods (D 1 - D 5).

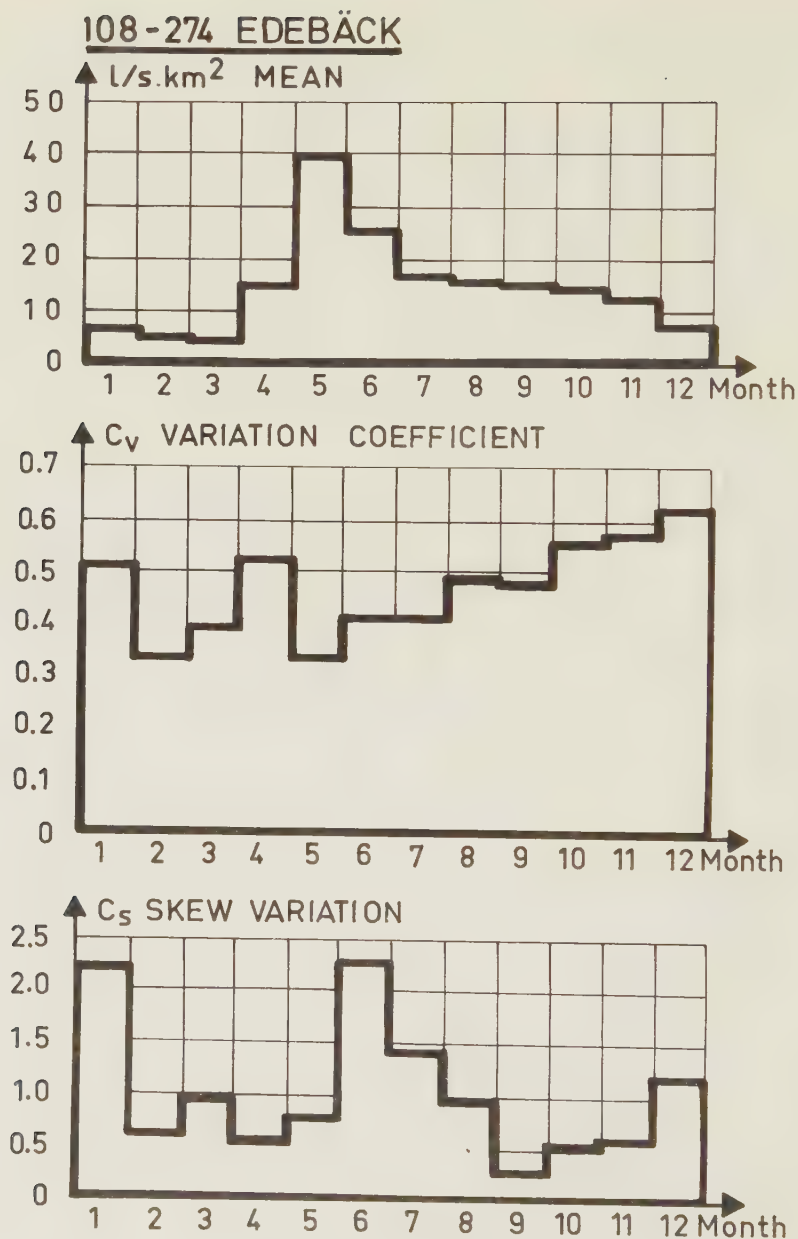


Fig 3.12 Graphs of monthly means and coefficients of variation and skewness for region D.

TABLE 3.11 Monthly mean runoff values for region D

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
42-358	3.7	2.9	2.2	4.4	52.8	55.7	27.1	19.8	18.3	16.9	9.5	5.9
42-826	3.2	2.2	2.2	5.4	65.7	72.2	35.6	22.6	22.6	20.5	10.8	6.5
42-829	4.6	3.5	3.1	5.8	48.2	43.6	23.1	18.1	17.0	15.0	10.4	6.9
48-106	5.2	4.7	4.5	10.1	43.7	28.0	17.3	16.1	15.7	13.3	9.7	6.5
48-107	5.2	4.7	4.4	9.0	35.9	23.2	14.4	13.9	13.4	11.5	9.0	6.6
53-121	7.4	6.2	5.5	9.1	28.3	26.2	16.2	13.9	14.3	13.9	13.1	9.6
108-	5.9	5.0	4.5	8.6	40.3	33.3	22.2	19.7	18.3	16.1	12.2	8.1
108-274	6.1	4.8	4.7	15.2	40.2	25.9	17.1	16.1	15.8	15.4	13.2	8.5

TABLE 3.12 Monthly variation coefficients for region D

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
42-358	0.49	0.62	0.57	0.80	0.33	0.36	0.43	0.44	0.49	0.40	0.39	0.38
42-826	0.41	0.48	0.42	0.83	0.34	0.36	0.43	0.43	0.55	0.45	0.41	0.55
42-829	0.34	0.40	0.40	0.72	0.33	0.40	0.40	0.45	0.49	0.39	0.31	0.29
48-106	0.42	0.49	0.52	0.70	0.32	0.45	0.44	0.44	0.50	0.41	0.41	0.39
48-107	0.36	0.38	0.41	0.55	0.38	0.47	0.41	0.48	0.55	0.44	0.43	0.37
53-121	0.40	0.32	0.28	0.39	0.37	0.37	0.35	0.47	0.50	0.47	0.41	0.35
108-	0.26	0.21	0.18	0.62	0.33	0.37	0.36	0.39	0.41	0.47	0.42	0.33
108-274	0.51	0.33	0.39	0.52	0.33	0.41	0.41	0.49	0.48	0.56	0.57	0.62

TABLE 3.13 Monthly skew coefficients for region D

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
42-358	3.33	3.59	3.56	1.13-0.09	0.49	1.60	0.43	0.85	0.39	1.55	1.18	
42-826	1.21	3.93	2.26	1.09-0.32	0.48	1.41	0.41	1.30	0.44	1.02	1.50	
42-829	1.15	1.63	1.14	1.10	0.30	0.77	1.23	0.94	0.82	0.45	0.84	0.48
48-106	1.37	1.82	1.79	1.62	0.52	1.60	1.22	0.68	1.00	0.58	0.47	0.48
48-107	0.82	1.29	1.62	1.07	0.80	1.97	1.18	0.81	1.15	0.63	0.88	0.35
53-121	2.18	1.83	0.68	0.90	0.76	1.62	0.65	1.08	0.90	0.89	0.61	0.09
108-	0.62	0.35	0.01	2.17	0.00	1.29	1.30	0.84	0.07	0.60	0.36	0.81
108-274	2.17	0.57	0.99	0.51	0.75	2.24	1.39	0.95	0.25	0.54	0.60	1.18

3.3.1.5 Forest and Coastal Rivers in Norrland

The coefficients calculated from the two series of this group of rivers are given in table 3.14 - 16 and for one of them these are also represented in a graph Fig 3.13. The region is not marked on the map as it coincides with the costal areas of regions A, B and C. The distinctions that can be made from these, is that the variation coefficient is high both in autumn and in winter. Further, the peaks in the mean values and variation coefficients are less expressed. The periods are the same five as before (E 1 - E 5).

TABLE 3.14 Monthly mean runoff values for region E

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
36-1101	4.1	3.4	3.4	11.7	31.0	18.6	9.6	8.3	9.6	10.3	9.6	7.6
36-1133	3.8	2.6	3.8	14.2	36.1	21.9	10.3	7.7	9.0	10.3	9.0	7.7

36-1101 MELLANSEL

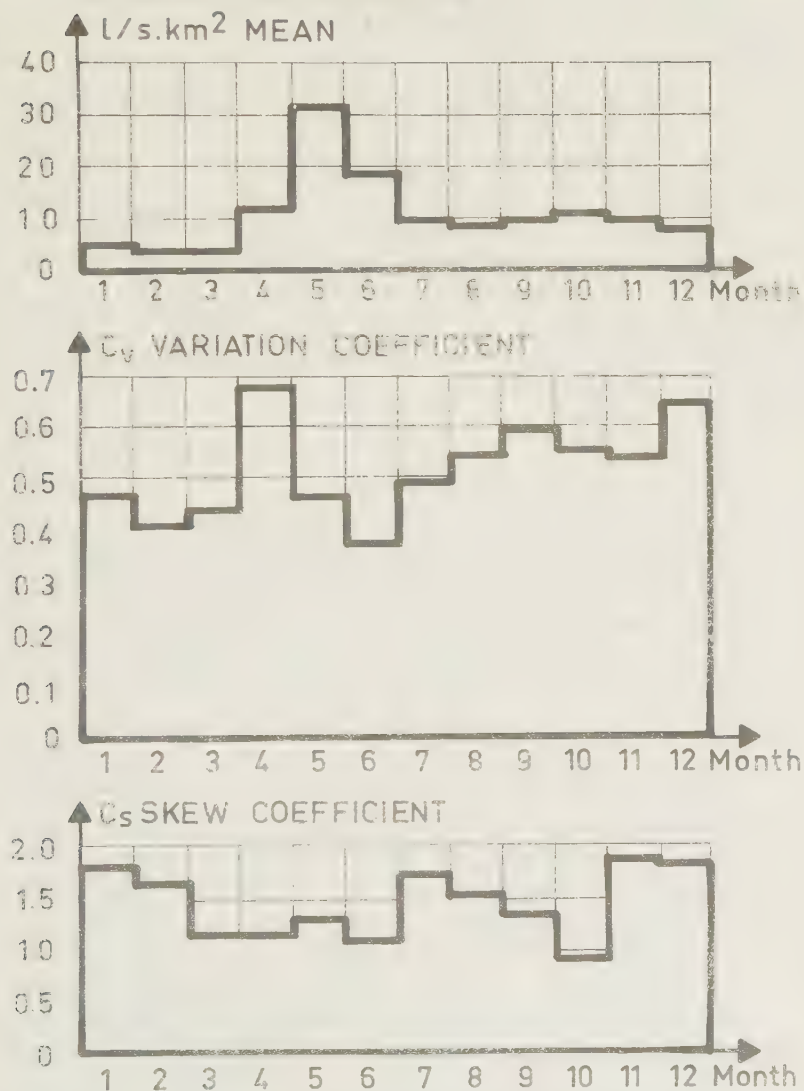


Fig 3.13 Graphs of monthly means and coefficients of variation and skewness for region E.

TABLE 3.15 Monthly variation coefficients for region E

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
36-1101	0.46	0.40	0.43	0.67	0.46	0.37	0.49	0.54	0.59	0.55	0.54	0.64
36-1133	0.55	0.42	0.55	0.62	0.42	0.38	0.57	0.69	0.75	0.65	0.68	0.90

TABLE 3.16 Monthly skew coefficients for region E

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
36 1101	1.78	1.61	1.13	1.14	1.27	1.06	1.70	1.50	1.31	0.87	1.82	1.78
36-1133	1.98	1.52	2.04	0.36	0.74	0.63	1.16	1.04	1.26	0.73	1.26	1.85

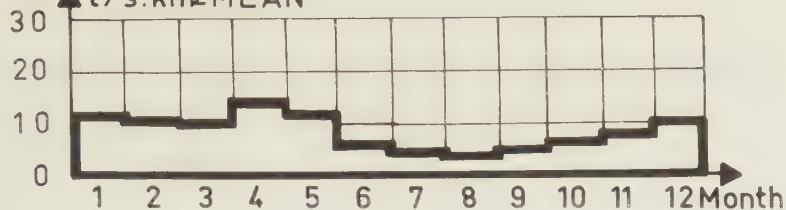
3.3.1.6 Rivers of the South of Sweden

The parameters for the two subgroups of rivers from this region are given in tables 3.17 - 22 and one example from the east subregion (Fig 3.14) and one from the west subregion (Fig 3.15) are shown. While in the regions A - E snowmelt is the main runoff forming factor, in the south of Sweden it is replaced by rain or combination of rain and snowmelt. Melin (1970) has observed an intermediate group of rivers between these two main groups (H-Rivers with mixed regime) which has not been considered here. They differ from the rivers of the south of Sweden by their higher spring floods. The change from one regime to another, however, is not very abrupt. We have noticed a clear tendency of an increase of the variation coefficient, especially in autumn, with the movement southwards. A corresponding peak in the mean has developed at this time.

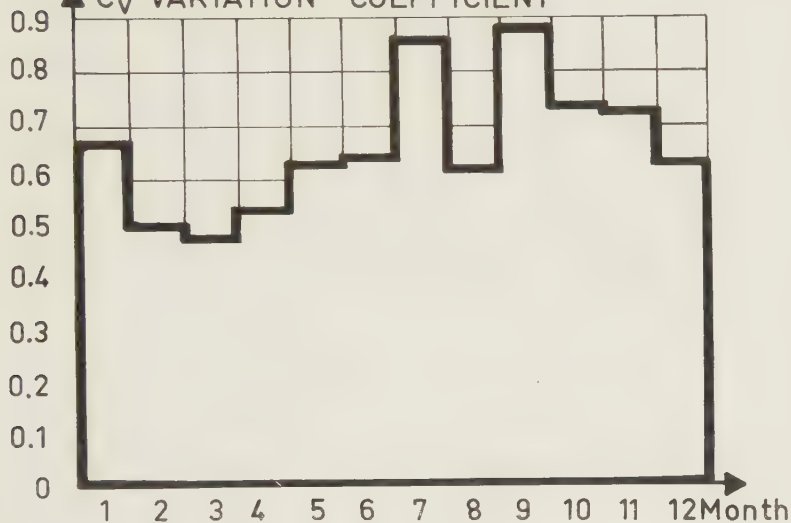
The southern region has been divided by Melin into one eastern and one western subregion, due to the differences in the runoff volume. One can also see this on the map Fig 3.1 by Kuprijanov. The eastern subregion is relatively dry, which we can see comparing the average values from the two subregions. This is expressed best of all during

74-177 JÄRNFORSÉN

▲ l/s.km²MEAN



▲ C_v VARIATION COEFFICIENT



▲ C_s SKEW COEFFICIENT

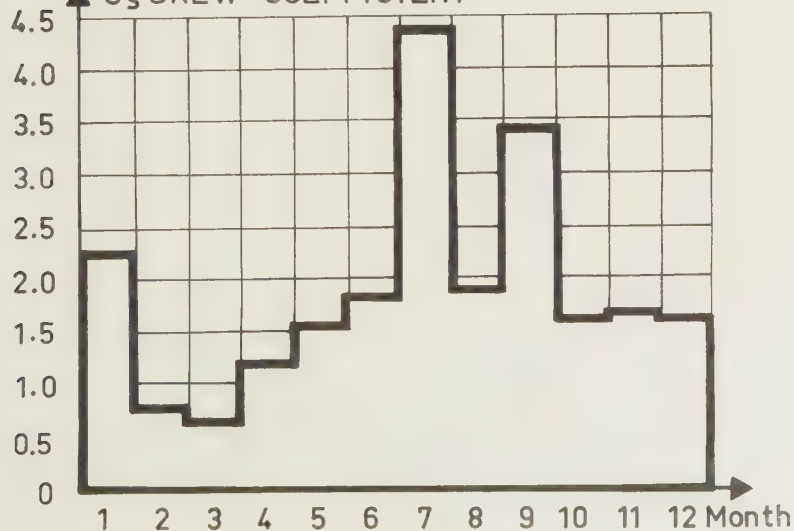


Fig 3.14 Graphs of monthly means and coefficients of variation and skewness for region Fa.

summer and autumn, when for the eastern part the mean is very low. At the same time we have very high values in the variation coefficient. The main pattern is the same for the western rivers, though not that extreme. Further difference between the two subregions can be found during late autumn and early winter, when for most southern and western rivers we can observe an additional high water period besides the spring flood. This high water period, as well as the increase in the runoff for the eastern rivers is, of course, the result of precipitation in connection with low evapotranspiration and filled up ground water storages. As far as the skew coefficient is concerned, it shows as before a somewhat irregular pattern. The tendency that high values of this parameter follow high values in the variation coefficient is apparent. The general picture is not as homogeneous as for the previous regions. We distinguish the following three periods, for both eastern and western subregions:

- Fa 1 and Fb 1: Winterflow; moderate mean and moderate variation coefficient.
- Fa 2 and Fa 2: Spring flow; moderate mean and moderate variation coefficient.
- Fa 3 and Fa 3: Summer and Autumn low water period; low mean value and very high variation coefficient.

TABLE 3.17 Monthly mean runoff values for region Fa

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
65-148	7.5	7.5	7.5	9.2	11.2	8.7	5.6	4.2	3.9	3.9	5.3	7.0
74-177	10.5	10.0	9.5	13.7	11.1	5.8	4.2	3.7	4.7	5.8	7.9	9.5
74-178	8.9	8.9	9.1	13.2	11.5	5.5	4.1	3.4	3.8	4.1	7.4	8.6
75-855	9.0	9.7	10.4	14.9	14.2	5.2	2.2	2.2	3.0	3.7	6.7	9.7
88-1069	11.9	12.9	11.9	11.9	11.9	7.9	5.9	5.0	5.0	5.9	7.9	10.9
88-187	15.7	15.7	13.9	14.4	12.0	7.9	6.0	5.6	6.9	7.9	12.0	14.4

TABLE 3.18 Monthly variation coefficients for region Fa

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
65-148	0.65	0.49	0.36	0.38	0.48	0.56	0.55	0.58	0.56	0.57	0.64	0.66
74-177	0.66	0.50	0.48	0.53	0.62	0.63	0.86	0.61	0.88	0.73	0.72	0.62
74-178	0.47	0.45	0.44	0.46	0.72	0.68	0.89	0.78	1.12	0.71	0.86	0.73
75-855	0.59	0.60	0.65	0.59	0.70	0.56	0.72	1.24	1.24	1.08	1.11	0.87
88-1069	0.46	0.45	0.49	0.56	0.54	0.50	0.44	0.45	0.86	0.78	0.75	0.66
88-187	0.43	0.49	0.48	0.34	0.45	0.40	0.43	0.39	0.60	0.55	0.63	0.56

TABLE 3.19 Monthly skew coefficients for region Fa

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
65-148	1.25	0.45	-0.20	0.62	1.05	1.69	1.54	1.59	1.14	0.62	0.83	0.88
74-177	2.45	0.76	0.65	1.19	1.54	1.81	4.39	1.89	3.47	1.63	1.67	1.63
74-178	0.66	0.43	0.96	1.07	1.92	1.66	2.51	2.02	3.82	1.23	1.95	1.46
75-855	1.56	1.27	1.39	1.49	1.27	1.02	2.08	4.00	2.55	1.78	1.80	1.30
88-1069	0.51	0.48	0.90	1.94	1.55	1.34	1.15	0.96	2.86	1.65	1.40	1.17
88-187	0.89	0.44	1.17	0.81	2.19	1.35	1.57	0.70	1.69	0.74	1.65	1.93

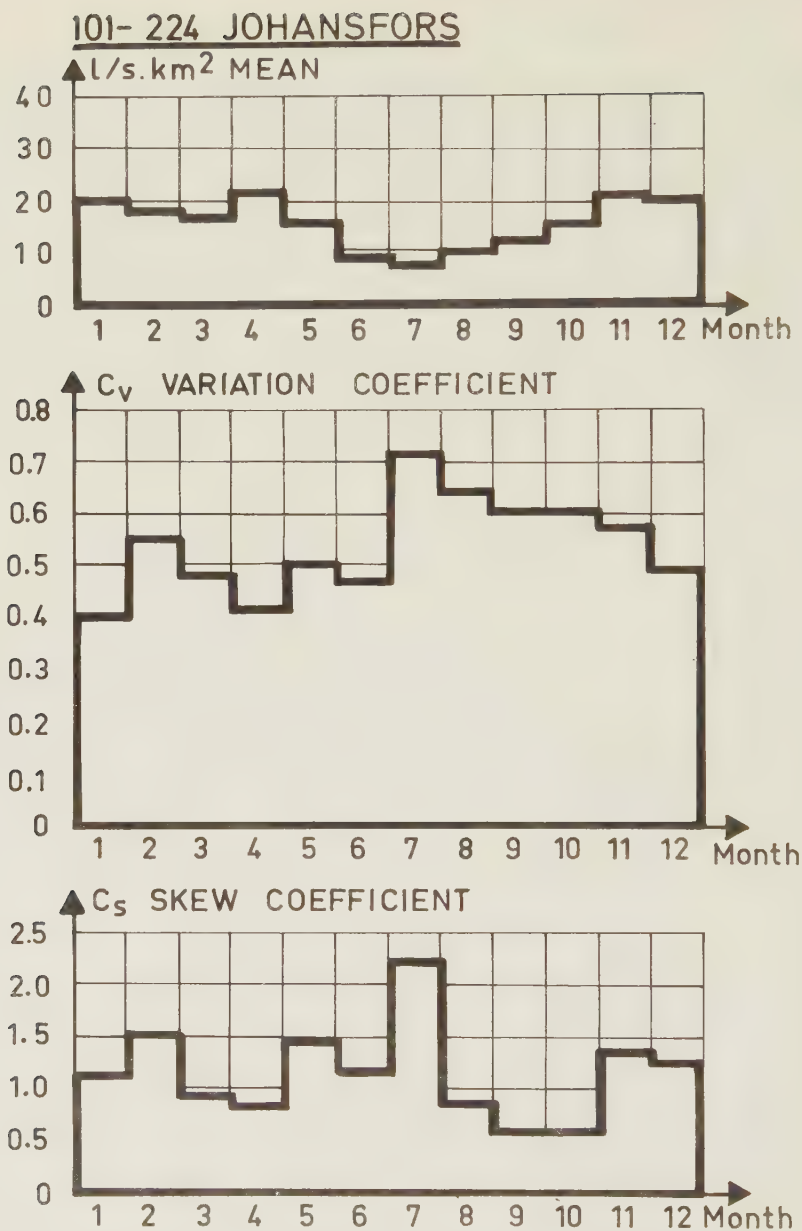


Fig 3.15 Graphs of monthly means and coefficients of variation and skewness for region Fb.

TABLE 3.20 Monthly mean runoff values for region Fb

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
96-1008	15.8	16.8	14.7	15.8	8.4	5.3	5.3	5.3	6.3	8.4	11.6	14.7
98-219	14.6	15.0	14.4	15.1	13.1	8.9	7.5	7.3	8.2	9.3	13.0	14.6
98-1048	17.6	17.0	16.4	16.4	15.2	10.9	7.9	6.7	6.7	7.9	12.1	15.2
101-224	19.8	17.8	17.0	21.1	15.4	8.5	7.7	10.1	11.7	15.0	20.2	19.8
101-923	17.6	15.7	14.5	17.6	14.5	7.9	7.9	8.5	10.6	14.5	20.6	15.7
103-974	20.6	19.1	17.1	19.5	15.9	9.5	7.9	8.7	10.7	15.1	21.4	20.6
105-227	23.6	20.4	15.3	16.7	12.5	6.5	6.5	8.8	12.0	17.1	22.2	21.8

TABLE 3.21 Monthly variation coefficients for region Fb

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
96-1008	0.46	0.58	0.49	0.57	0.50	0.31	0.47	0.70	0.63	0.64	0.60	0.49
98-219	0.37	0.43	0.41	0.32	0.44	0.44	0.52	0.42	0.56	0.48	0.54	0.46
98-1048	0.40	0.35	0.39	0.32	0.39	0.50	0.58	0.59	0.72	0.65	0.70	0.58
101-224	0.40	0.55	0.48	0.41	0.50	0.47	0.71	0.64	0.60	0.60	0.57	0.49
101-923	0.43	0.51	0.55	0.41	0.68	0.61	0.89	0.62	0.58	0.60	0.61	0.45
103-974	0.43	0.50	0.52	0.36	0.54	0.58	0.81	0.71	0.73	0.65	0.59	0.52
105-227	0.48	0.68	0.62	0.43	0.74	0.60	0.80	0.66	0.74	0.73	0.60	0.56

TABLE 3.22 Monthly skew coefficients for region Fb

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
96-1008	0.35	0.61	0.90	1.40	1.40	0.42	1.09	2.37	1.37	1.09	1.24	0.58
98-219	0.86	0.81	0.72	0.26	1.77	2.23	3.17	0.84	1.58	0.74	1.61	1.34
98-1048	0.30-0.01		0.54	0.41	1.45	2.08	1.76	1.06	0.95	0.64	1.16	0.71
101-224	1.10	1.48	0.92	0.83	1.45	1.16	2.36	0.85	0.60	0.60	1.33	1.24
101-923	0.52	0.45	1.86	0.94	1.87	1.27	2.94	0.75	0.98	0.74	1.14	0.69
103-974	0.59	0.71	0.84	0.44	1.31	1.21	3.44	1.14	1.12	0.77	0.74	0.97
105-227	0.72	1.89	1.20	0.81	2.70	1.94	3.48	1.10	1.32	1.16	0.73	1.04

3.3.1.7 Rivers from the large Swedish lakes

The parameters are given in table 3.23 - 25 and one example is drawn in Fig 3.16. The monthly variations in the empirical moments for the outflows from the large Swedish lakes are small. Both the variation coefficient and the skew coefficient have small absolute values. An annual periodicity can be observed. In particular, the variations in the skew coefficient at 108-243 Sjötorp are remarkable, because of the coefficient's negative values during summer and autumn. They are the only negative values of this coefficient observed for all the series.

TABLE 3.23 Monthly mean runoff values for region G

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
67-154	6.0	6.0	6.0	6.6	7.2	7.7	7.9	7.7	7.4	6.8	6.3	6.1
108-243	11.7	11.7	11.5	11.3	11.4	11.4	11.2	11.3	11.4	11.5	11.7	11.7

TABLE 3.24 Monthly variation coefficients for region G

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
67-154	0.31	0.33	0.32	0.30	0.27	0.24	0.22	0.22	0.24	0.26	0.28	0.30
108-243	0.25	0.26	0.26	0.26	0.26	0.30	0.30	0.25	0.23	0.23	0.24	0.26

TABLE 3.25 Monthly skew coefficients for region G

Station	Month											
	Jan	Feb	March	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
67-154	0.45	0.50	0.66	0.66	0.63	0.53	0.57	0.53	0.46	0.35	0.37	0.32
108-243	0.29	0.35	0.54	0.43	0.02	-0.61	-0.76	-0.64	-0.24	-0.05	0.39	0.43

67-154 MOTALA

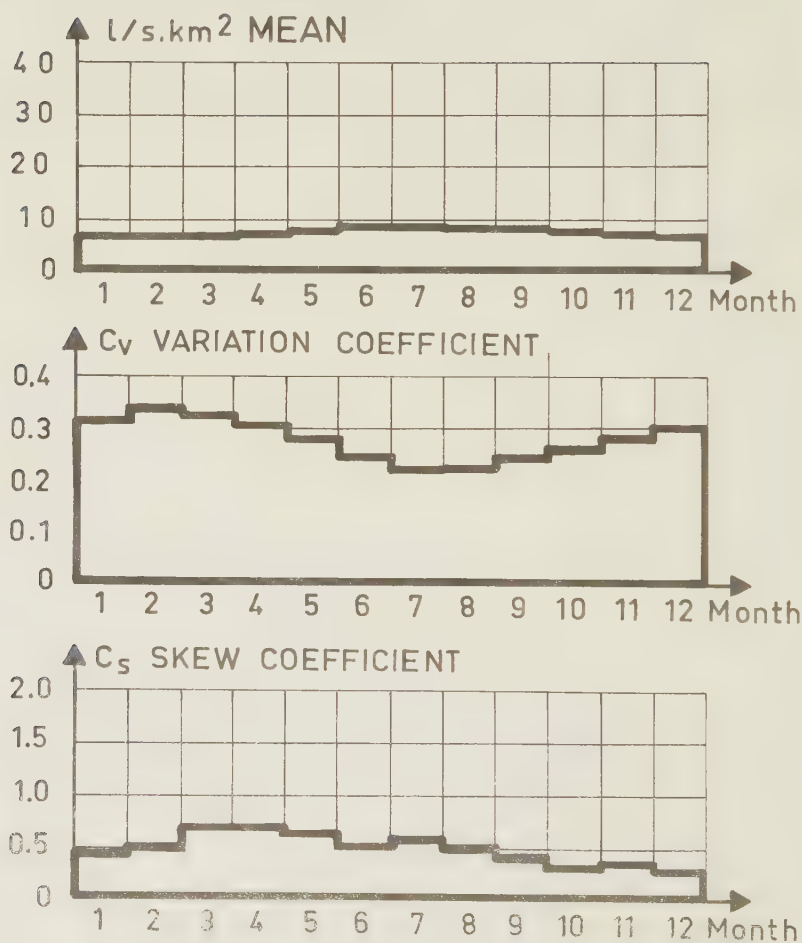


Fig 3.16 Graphs of monthly means and coefficients of variation and skewness for region G.

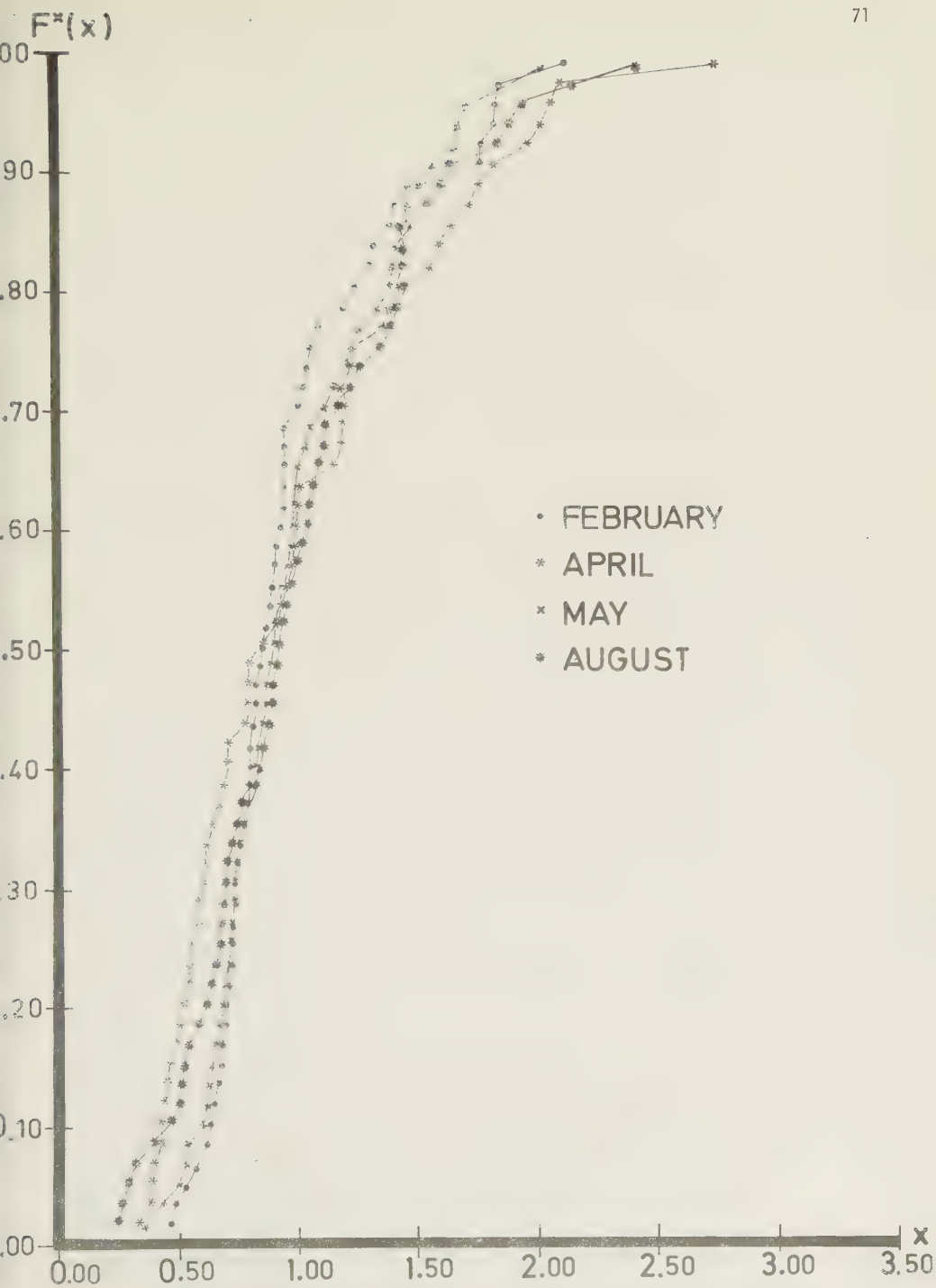


Fig 3.17 Plotted empirical distributions of monthly river runoff
for some months at 48-107 Ljusdal (northern Sweden).

3.3.2 General conclusions

The review of some general patterns for the regional differences in the empirical moments for monthly runoff given above, could be based, of course, on the plotted empirical distributions. They are, however, a weaker instrument, which is more difficult to interpret. The peculiarities of the different regions and the different periods are though seen clearly. In Figs 3.17 and 3.18 two examples are shown, one from the north of Sweden (48-107 Ljusdal) and one from the south (74-177 Järnforsen). The specific periods, described above can be clearly distinguished.

The given examples of the empirical moments show regularities caused mainly by the climatic runoff forming factors. Within each region the pattern of the variations of the mean and the variation coefficient are the same, though we have differences in the specific parameters. Two main regions can be formed by the regions A-E and F, between which we have differences not only in the specific parameters but also in the patterns of their variations.

In table 3.26 the range of the parameters for each region and each period are shown. Part of the differences can be explained by statistical errors. This concerns especially the skew coefficient. The sudden jumps of this coefficient and relatively large irregularities are probably caused by them. To some extent they can account also for the irregularities in the behaviour of the variation coefficient.

Differences are also caused by variations in the physiographic conditions within each region, and the specific characteristics of the catchment.

Attempts to find relations between certain physiographic factors and runoff parameters with the help of regression analysis, have been done for Scandinavian conditions (Mustonen 1967, Nordseth 1973, Gottschalk and Falk 1973). The two first studies deal with annual values for the whole Finland and Norway, respectively. Nordseth uses only morphometric factors in his study, while Mustonen includes also climatic ones and thus can explain more of variance in the runoff

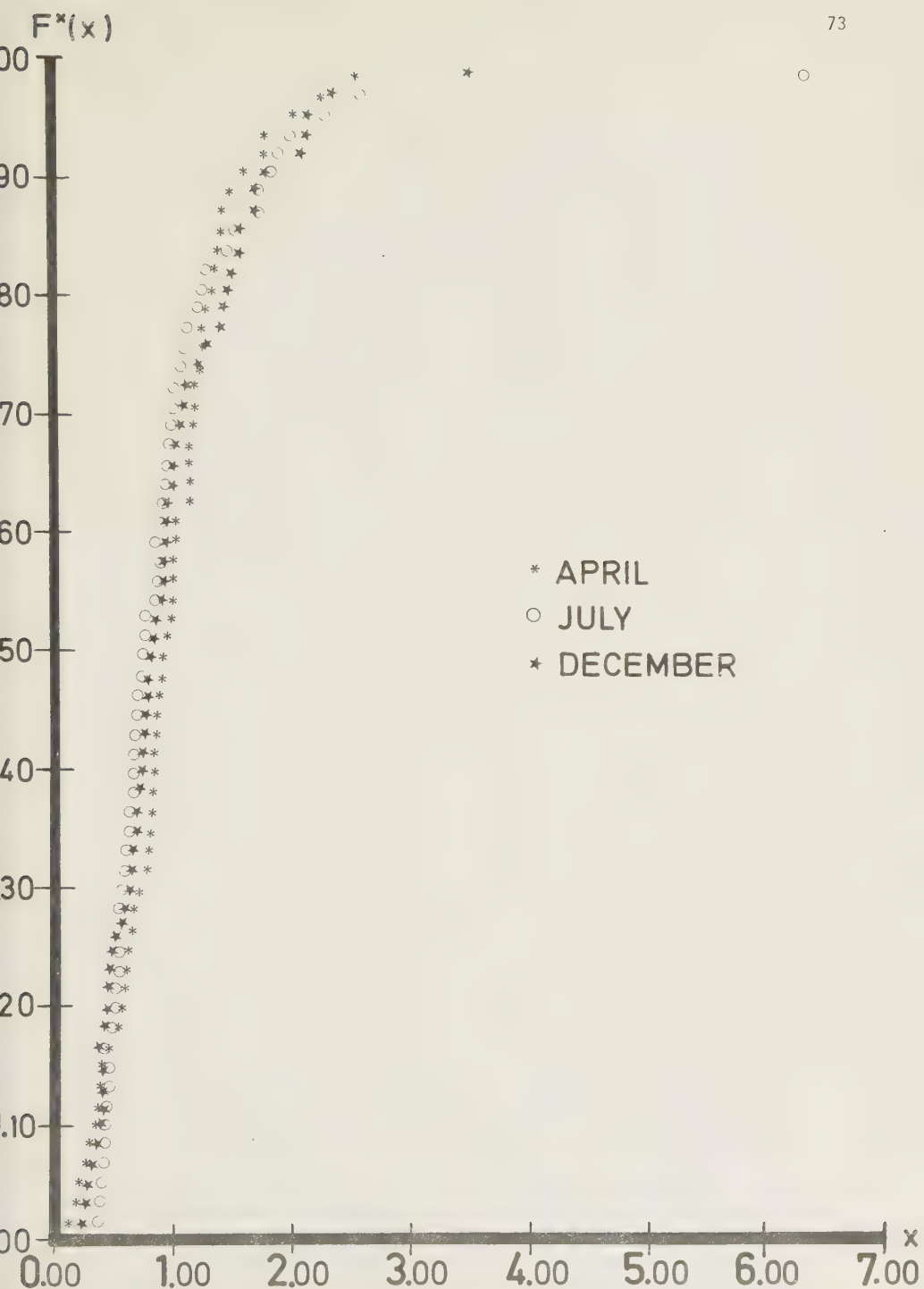


Fig 3.18 Plotted empirical distributions of monthly river runoff for some months at 74-177 Järnforsen (southern Sweden).

TABLE 3.26 Range of monthly mean values, variation and skew coefficients of runoff

Parameter Period	MV	C_v	C_s
A_1	3-12	0.18 - 0.46	0.3 - 2.65
A_2	3-12	0.50 - 0.59	0.4 - 2.60
A_3	29-60	0.20 - 0.40	0.15 - 1.25
A_4	14-46	0.18 - 0.40	0.00 - 1.80
A_5	10-23	0.33 - 0.53	0.30 - 2.05
B_1	2-13	0.20 - 0.41	0.30 - 3.50
B_2	18-28	0.45 - 0.77	0.40 - 1.30
B_3	38-90	0.30 - 0.34	0.20 - 1.45
B_4	17-53	0.22 - 0.47	0.25 - 0.70
B_5	14-28	0.30 - 0.43	0.60 - 1.15
C_1	3-14	0.29 - 0.50	0.05 - 1.30
C_2	5-6	0.56 - 0.65	1.70 - 2.00
C_3	33-73	0.27 - 0.46	0.10 - 1.10
C_4	13-34	0.34 - 0.64	0.45 - 3.10
C_5	14-23	0.48 - 0.62	-0.20 - 1.05
D_1	2-13	0.28 - 0.62	0.00 - 3.55
D_2	5-10	0.39 - 0.83	0.50 - 2.20
D_3	26-72	0.32 - 0.40	-0.30 - 1.60
D_4	14-35	0.35 - 0.49	0.40 - 2.24
D_5	12-22	0.40 - 0.55	0.05 - 1.30
E_1	2-8	0.40 - 0.90	1.20 - 2.00
E_2	12-14	0.62 - 0.67	0.35 - 1.10
E_3	31-36	0.42 - 0.46	0.75 - 1.25
E_4	8-22	0.38 - 0.69	0.65 - 1.70
E_5	9-10	0.54 - 0.75	0.75 - 1.80
Fa_1	5-15	0.36 - 0.73	-0.20 - 1.65
Fa_2	9-15	0.34 - 0.72	0.80 - 2.15
Fa_3	2-9	0.40 - 1.12	0.60 - 4.40

Table 3.26

Parameter Period	MV	C_v	C_s
Fb ₁	15-23	0.35 - 0.68	0.00 - 1.35
Fb ₂	13-21	0.32 - 0.74	0.25 - 2.70
Fb ₃	5-15	0.10 - 0.90	0.40 - 3.50
G	6-12	0.22 - 0.33	-0.76 - 0.66

values. Mustonen's conclusions are that the found regression equations can serve as a good working instrument, but the physical interpretation of the results is a very difficult task. Gottschalk and Falk deal with monthly runoff values within a region with small variations in geology and climate (Själland and Scania). As independent variable they use, first of all, basin characteristics and the only climatic factor is monthly precipitation. The explained variance is 73-84 % of the total variance in the mean and 61-82 % in the variation coefficient of monthly runoff.

Maps over annual means (for references see Melin, 1970) and annual variation coefficients (Kuprijanov, 1960 and Tollan, 1972) have been plotted for Scandinavian conditions. As was discussed above, these coefficients are dependent on basin characteristics, and it seems difficult to include this factor when making a map. Example of maps over monthly coefficients is given by Kapitanova (1971).

It is a very important problem in hydrology to extrapolate information from areas with long observation series to others, where few or no observations are performed. Regression studies, like those sited above, show an approach to this problem, that is better than plotting of maps. For the application of the technique, presented further in this work, it is necessary to develop for areas with no observation such kind of relations. Much can be thereby gained by deviding of a large territory like Sweden, into regions which are relatively homogeneous as far as climate and geology are concerned. The variance in runoff parameters is reduced and a relation can be derived to explain the remaining variance within the region.

3.4 Analytical distribution functions

Fundamental in hydrologic practice are the normal distribution function and the gamma distribution function. Almost all continuous distribution functions applied to runoff phenomena can be derived from these. The wellknown expression for the normal distribution function is:

$$F(x; m, \sigma) = 1 / (\sigma \sqrt{2\pi}) \int_{-\infty}^x \exp(-(t-m)^2 / 2\sigma^2) dt \quad (3.10)$$

and the corresponding normal density function is:

$$f(x; m, \sigma) = 1 / (\sigma \sqrt{2\pi}) \exp(-(x-m)^2 / 2\sigma^2) \quad (3.11)$$

with the parameters m and σ . Setting $m = 0$ and $\sigma = 1$, we get the unit normal distribution defined by:

$$\phi(x) = 1 / \sqrt{2\pi} \int_{-\infty}^x \exp(-t^2 / 2) dt \quad (3.12)$$

$$\phi'(x) = 1 / \sqrt{2\pi} \exp(-x^2 / 2) \quad (3.13)$$

For further properties of the normal distribution see, for instance, Cramer (1948).

The one parameter gamma distribution function is defined by the following expression:

$$F(x; \gamma) = 1 / \Gamma(\gamma) \int_0^x t^{\gamma-1} \exp(-t) dt \quad (3.14)$$

where $\Gamma(\gamma)$ is the gamma function of the shape parameter γ . The expected value, variance and skew coefficient for this distribution are respectively

$$E(X) = \gamma, \text{Var}(X) = \gamma, C_s = 2 / \sqrt{\gamma}$$

These basic distributions are usually not applicable directly to runoff phenomena. It is, for instance, obvious that the normal distribution does not give a proper description of runoff as it is symmetric and has a range from $-\infty$ to ∞ , while runoff is always greater or equal

to zero and has a skewed distribution. The one parameter gamma distribution has these properties, but as it involves only one parameter, it is not flexible enough to describe the more complex distribution of runoff. This distribution gives, for instance, a constant relation between the variation coefficient and the skew coefficient $C_s = 2 C_v$. From the coefficients calculated in the previous part this relation varies in the interval 1.0 - 7.0. New probability functions that have this flexibility and own the above mentioned properties

are derived by replacing the random variable X by some function of it $u(X)$. In principle, the function u can be an arbitrary continuous differentiable function, but for practical application it is better to use monotonous increasing functions. The character of the function u depends on the basic distribution function, the character of the investigated phenomena and the available observation material. $u(X)$ must satisfy the following condition: if the density function differs from zero in the interval $u_1 < u < u_2$ and the considered random variable is defined in the interval $x_1 < X < x_2$ (for runoff process $x_1 = 0$ and $x_2 = \infty$, usually), the transformation function must satisfy $u(x_1) = u_1$ and $u(x_2) = u_2$.

The transformation can be done in two ways:

1. The new random variable $u(X)$ replaces X in the basic cumulative distribution function

$$P(X < x) = F(x) = \int_{-\infty}^x f(z) dz$$

We derive then the following expression for the cumulative distribution function of the variable X :

$$P(X < x) = \Psi(x) = F(u(x)) = \int_{-\infty}^{u(x)} f(z) dz \quad (3.15)$$

and for the density function we have:

$$\Psi(x) = \frac{d(\Psi(x))}{dx} = F'(u(x)) \frac{du(x)}{dx} = f(u(x)) \frac{du(x)}{dx} \quad (3.16)$$

2. The variable $u(x)$ replaces x in the basic density function $f(x)$. The expression for the density function then is:

$$\Psi(x) = c \cdot f(u(x)) \quad (3.17)$$

where c is a constant, which can be calculated from the expression:

$$\int_{-\infty}^{\infty} c f(u(x)) dx = 1$$

In general, the first method is the most convenient one.

3.4.1 Transformations of the normal distribution

The normal distribution can not be used directly to answer the demand to describe river runoff.

Instead, we can apply the method of translation, i.e. transforming variables in such a way that they may be considered to have a normal distribution. Two ways of approaching this problem can be found in literature. The first one is to suggest the analytical form of the transformation. From equations (3.15) and (3.16) the exact expressions for the distribution function and the density function can be then derived. A general system of distribution functions, based on transformations of the unit normal distribution has been suggested by Johnson (1949a) as an alternative to the system of K Pearson and Charlier for skew frequency curves. Johnson introduces four parameters as in the case of the Pearson system and gives the following general expression for the transformation function:

$$u = \gamma + \delta z \left(\frac{x - \xi}{\lambda} \right) = \gamma + \delta z(y)$$

where γ , δ , ξ and λ are parameters and z a simple function. Three special systems are developed with $z(y) = \ln y$, $z(y) = \ln \{y / (1-y)\}$ and $z(y) = \ln \{y + \sqrt{(y^2+1)}\}$, respectively. In future we shall consider only transformations that have been applied in hydrology, and for further references see Johnson (1949a), Aitchison and Brown (1957).

The most well known transformations are:

$$u(x) = \ln(x) \tag{3.18}$$

and

$$u(x) = \ln(x - \epsilon)$$

where ϵ in the last expression represents a lower bound. Assuming $u(x)$ to be normally distributed (eq 3.10) we get a two parameter and three parameter log normal distribution respectively:

$$f(x; m_n, \sigma_n) = 1 / (\sigma_n \sqrt{2\pi} x) \exp \{-(\ln x - m_n)^2 / 2 \sigma_n^2\} \quad (3.20)$$

$$f(x; m_n, \sigma_n, \epsilon) = 1 / (\sigma_n \sqrt{2\pi} (x - \epsilon)) \exp \{-(\ln(x - \epsilon) - m_n)^2 / 2 \sigma_n^2\} \quad (3.21)$$

These two distribution functions are widely applied to describe river runoff. Kartvelishvili (1967) suggests for the same purpose the following transformation for the unit normal distribution (eq 3.12):

$$u(x) = a \ln(\ln(1+x)) + b \ln(x) + c \sqrt{x} \quad (3.22)$$

which gives a three parameter distribution:

$$f(x; a, b, c) = 1 / \sqrt{2\pi} \exp \left\{ -\frac{1}{2} (a \ln(\ln(1+x)) + b \ln(x) + c \sqrt{x})^2 \right\} \cdot (a / (\ln(1+x) (1+x)) + b / x + c / (2 \sqrt{x})) \quad (3.23)$$

Other transformations to be mentioned are an m -th order polynomial (Yevjevich, 1972a) and a cube root and generally n -th root Stidd (1953, 1970), etc, more references can be found in literature (Karganova, 1971, Kazcmarek, 1970 and Yevjevich, 1972b).

The second approach to derive transformation functions is to start from some other distribution function that suits well the considered phenomena. Denoting this $F(x)$ and further using the unit normal distribution, (eq 3.12) we drive the following expression by inserting in eq 3.15:

$$F(x) = \Phi(u) \quad (3.24)$$

The relation is represented in Fig 3.19. Using eq 3.16 we in the same way find:

$$f(x) = \phi(u) \frac{du}{dx} = 0 \quad (3.25)$$

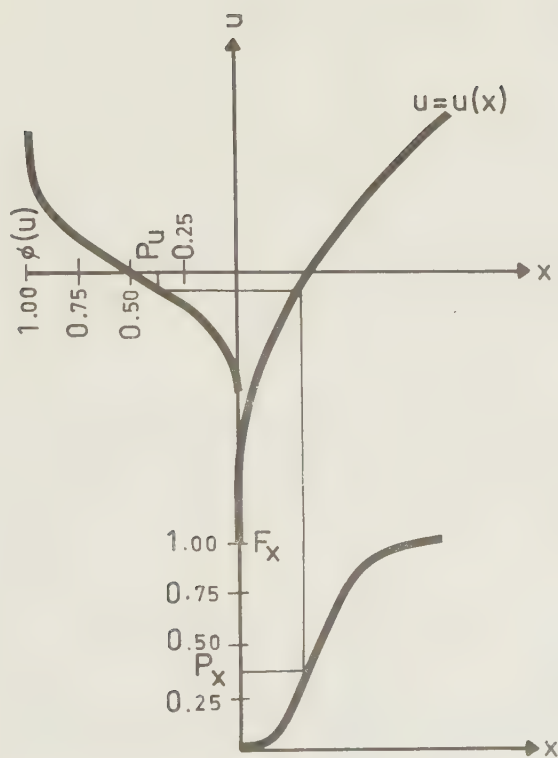


Fig 3.19 Graphical illustration of transformation of the unit normal distribution.

It is difficult to solve either of these equations analytically and in many cases impossible. Numerical solutions can, however, be obtained. Kazcmarek (1970) has suggested an approximate analytical expression, when we for $F(x)$ have the Pearson type III distribution. This distribution is actually the same as the three parameter gamma distribution eq (3.37). The transformation has the following expression:

$$u(x) = a_Y (1/\beta)^{0.28} (x - \epsilon)^{0.28} - b_Y$$

where a_Y and b_Y are parameters dependent on the parameter γ . Beard (1965) uses for the same purpose the equation:

$$u(x) = (2 / C_S) \{ (C_S (x - C_S / 6) / 6 + 1) - 1 \}$$

Computer algorithms to solve eq (3.24) for this case are referenced in an article by Borgman and Amorocho (1970). Moran (1969) suggests to use the two parameter gamma distribution (eq 3.34). For the solution of eq 3.24 an implicit numerical procedure is accepted which involves numerical integration of equations (3.12) and (3.34) on a computer to find corresponding values of (x, u) .

This latter approach usually gives complex transformation functions or complex numerical derivations. Further, there is no reason to believe that either of the distributions Pearson type III and two parameter gamma are physically more suitable than transformed normal distributions.

3.4.2 Empirical transformation functions of the unit normal distribution

Empirical transformation functions of the normal distribution can be derived with the help of eq (3.24) and the results from part 3.2.2. An equation to calculate the empirical function

$$p_i = \Phi(u) \tag{3.26}$$

where p_i is the i -th order statistic, corresponding to the i -th observed value x_i^1 , and $\Phi(u)$ is the unit normal distribution. Let

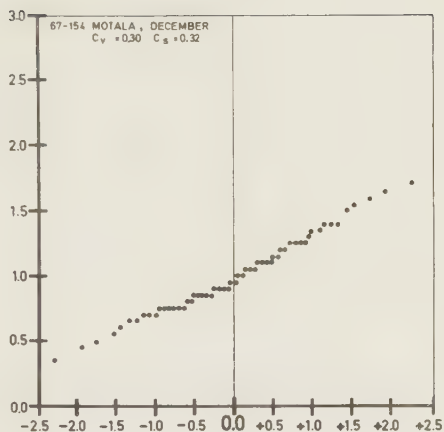
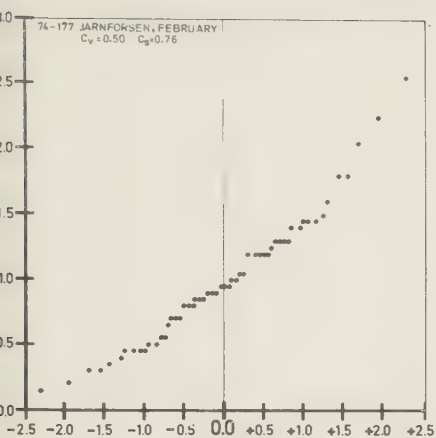


Fig 3.20 Empirical transformation functions of the unit normal distribution for monthly river runoff.

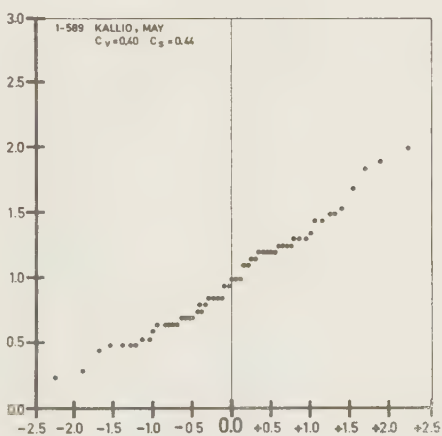
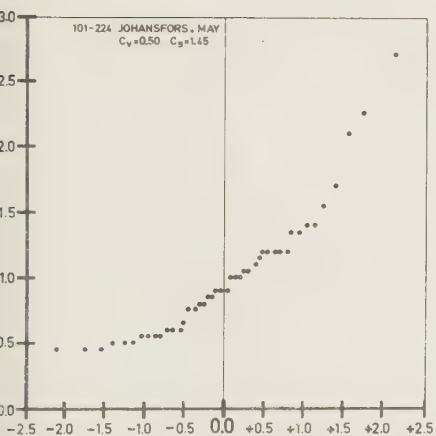


Fig 3.21 Empirical transformation functions of the unit normal distribution for monthly river runoff.

the empirical transformation function be given by sets of pairs (x^i, u^i) , $i = 1, \dots, N$, where N is the number of observations and u^i - the expected value of the transformed β -variable corresponding to p_i . A good estimate for u^i is (cf 3.2.2):

$$u^i = \Phi^{-1}(p'_i) = \Phi^{-1}((i - 3/8) / (N + 1/4)) \quad (3.27)$$

where Φ^{-1} is the inverse to the unit normal distribution. Numerically values of u^i can be found explicitly by using a formula after Hastings (1955):

$$u^i = t - \frac{e_0 + e_1 t + e_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}, \quad t = \sqrt{\ln(1 / p'_i)^2} \quad (3.28)$$

where $e_0, e_1, e_2, d_1, d_2, d_3$ are constants. If we use instead of p'_i eq 3.5 or eq 3.9 the differences in the result are not negligible especially for i close to zero or N . This is illustrated by table 3.27, where p_i and u^i have been calculated for N equal 41.

Empirical transformation functions were plotted for the observation series in table 3.1 for each month. The general shape of these transformation functions is, of course, to a very great extent described by the empirical moments. Thus follows the main pattern of the regional variations in these as have been already discribed in part 3.3.1. Examples of plotted transformation functions of the normal distribution function are given in Figs 3.20 - 25. For negative values of u the module x is close to zero and with the increase of u from negative to positive values x also increases. The slope of the transformation curve is described first of all by the variation coefficient (Fig 3.20). The transformation function usually does not represent a straight line but has a more complex form. In the lower end the function is convex with a lower bound at $x = 0$ or some positive value (Fig 3.21). In the upper part the empirical function can be both convex or concave (Fig 3.22 - 23).

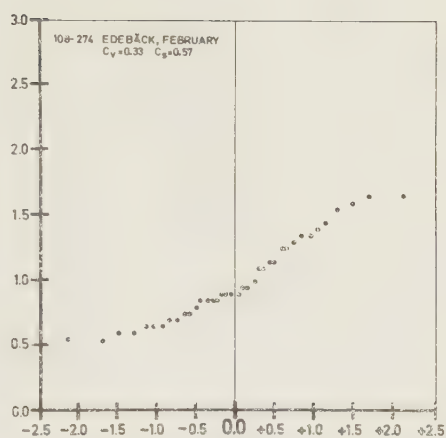
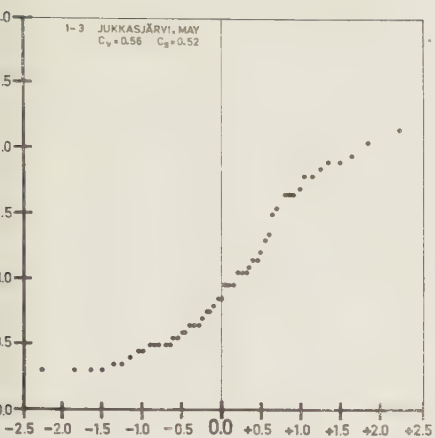


Fig 3.22 Empirical transformation functions of the unit normal distribution for monthly river runoff.

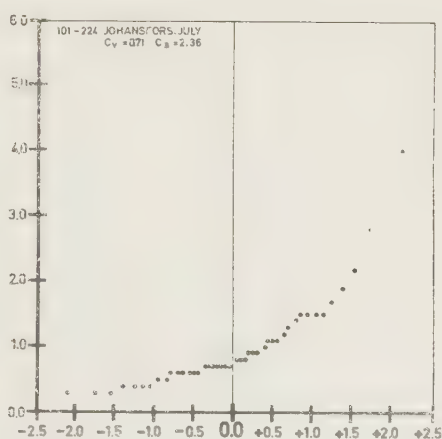
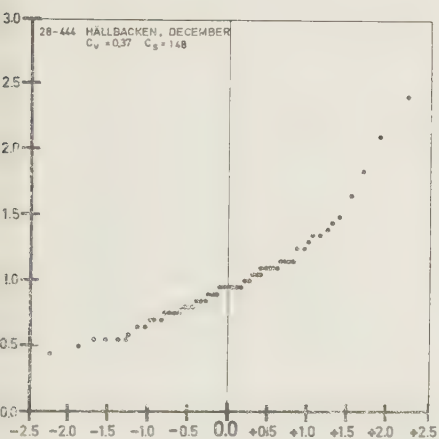


Fig 3.23 Empirical transformation functions of the unit normal distribution for monthly river runoff.

Table 3.27 Example of transformed B-values $u^i = \phi^{-1}(p_i)$ for $N = 41$

i	$p_i = \frac{i - 3/8}{N + 1/4}$	u^i	$p_i = \frac{i - 0.25}{N + 0.5}$	u^i	$p_i = \frac{i}{N + 1}$	u^i
41	0.985	2.166	0.982	2.095	0.976	1.981
40	0.961	1.758	0.958	1.726	0.952	1.669
39	0.936	1.525	0.934	1.504	0.929	1.465
38	0.912	1.354	0.910	1.339	0.905	1.310
37	0.888	1.215	0.886	1.206	0.881	1.180
36	0.864	1.096	0.862	1.086	0.857	1.068
35	0.839	0.992	0.837	0.982	0.833	0.967
34	0.815	0.897	0.813	0.889	0.810	0.876
33	0.791	0.810	0.789	0.803	0.786	0.792
32	0.767	0.728	0.765	0.722	0.762	0.712
31	0.742	0.651	0.741	0.646	0.738	0.637
30	0.718	0.577	0.717	0.574	0.714	0.566
29	0.694	0.507	0.693	0.504	0.690	0.497
28	0.670	0.439	0.669	0.437	0.667	0.431
27	0.645	0.373	0.645	0.372	0.643	0.366
26	0.621	0.309	0.620	0.305	0.619	0.303
25	0.597	0.245	0.596	0.243	0.595	0.241
24	0.573	0.184	0.572	0.181	0.571	0.180
23	0.548	0.122	0.548	0.121	0.548	0.120
22	0.524	0.061	0.524	0.060	0.524	0.060
21	0.500	0.000	0.500	0.000	0.500	0.000
20	0.476	-0.061	0.476	-0.060	0.476	-0.060
19	0.452	-0.122	0.452	-0.121	0.452	-0.120
18	0.427	-0.184	0.428	-0.181	0.429	-0.180
17	0.403	-0.245	0.404	-0.243	0.405	-0.241
16	0.379	-0.309	0.380	-0.305	0.381	-0.303
15	0.355	-0.373	0.355	-0.372	0.357	-0.366
14	0.330	-0.439	0.331	-0.437	0.333	-0.431
13	0.306	-0.507	0.307	-0.504	0.310	-0.497
12	0.282	-0.577	0.283	-0.574	0.286	-0.566
11	0.258	-0.651	0.259	-0.646	0.262	-0.637
10	0.233	-0.728	0.235	-0.722	0.236	-0.712
9	0.209	-0.810	0.211	-0.803	0.214	-0.792
8	0.185	-0.897	0.187	-0.889	0.190	-0.876
7	0.161	-0.992	0.163	-0.982	0.167	-0.967
6	0.136	-1.096	0.138	-1.086	0.143	-1.068
5	0.112	-1.215	0.114	-1.206	0.119	-1.180
4	0.088	-1.354	0.090	-1.339	0.095	-1.310
3	0.064	-1.525	0.066	-1.504	0.071	-1.465
2	0.039	-1.758	0.042	-1.726	0.048	-1.669
1	0.015	-2.166	0.018	-2.095	0.024	-1.981

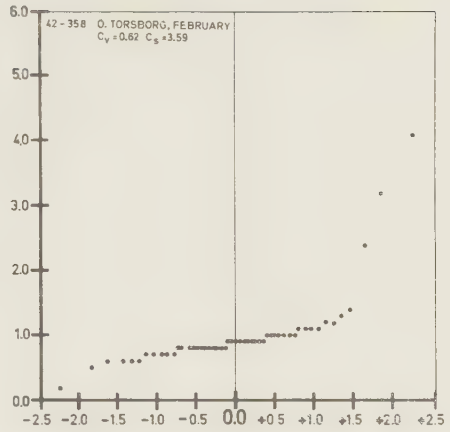
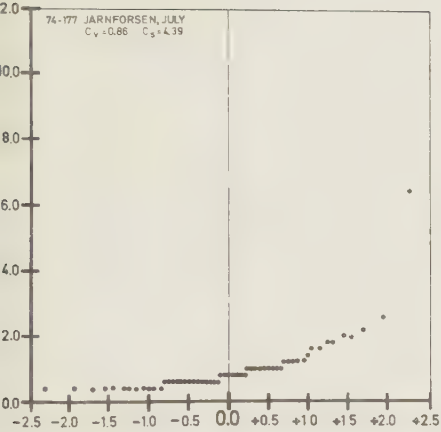


Fig 3.24 Empirical transformation functions of the unit normal distribution for monthly river runoff.

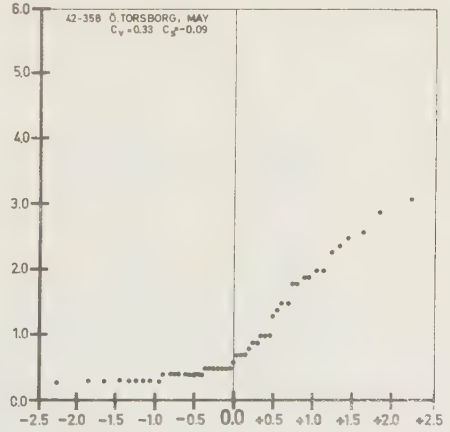
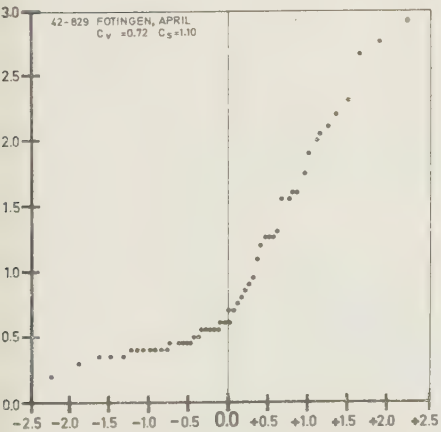


Fig 3.25 Empirical transformation functions of the unit normal distribution for monthly river runoff.

A convex form can indicate the existence of an upper bound. The deviation of the transformation function from a straight line is indicated by the skew coefficient (Fig 3.20, 23). Plots like these shown above give a good complement to the empirical moments for characterization of the observation series. In Fig 3.24 we can see the effect of very extreme situations. For the series 42-358 Ø Torsborg the high values are probably caused by mild winter periods with snowmelt. Another period, when the runoff can be of different origin, is the start of snowmelt, which usually gives a very peculiar form of the transformation curve with a distinct break point, Fig 3.25.

A very detailed description of the empirical curve can not be gained by the empirical moments. The statistical uncertainty, involved in the method for its construction, however, does not give any reason to extend the number of parameters for the description of the empirical function.

3.4.3 Analytical transformation functions of the unit normal distribution

The empirical transformation functions are defined only in the interval, where observations are given. Furthermore, they have usually sharp discontinuities that can be expected to be smoothed out if the number of observations is large. For extrapolation beyond the interval of observations and for smoothing out of the empirical function, analytical expressions for the transformation functions are used. The number of parameters in such analytical expressions should be quite few. As was discussed earlier, already the third moment gives rise to large uncertainties. A two parameter transformation should therefore be recommendable. For this study five two parameter transformations were tested and one transformation with three parameters. The choice of suitable analytical expressions for the transformation functions was guided by the form of the empirical functions. It was assumed that no upper or lower bounds exist.

The most simple two parameter transformation is that of a straight line, for which we write down the expression:

$$u(x) = a \cdot x - b \quad (3.29)$$

It neither answers well the genetical understanding of runoff nor fits very well the form of the empirical transformation functions. Nevertheless, it was accepted because of its simplicity.

A logarithmic expression describes well the form of the empirical functions. A two parameter logarithmic transformation function has the expression:

$$u(x) = a \ln(x) + b \quad (3.30)$$

As x tends to zero $u(x)$ should tend to minus infinity, and as x tends to infinity $u(x)$ should tend to infinity. The first condition can be provided by an expression of the form $u(x) = -b/x$. This expression tends to zero as x tends to infinity. We can add to this function another function that will be of negligible influence on u for small values of x and that tends to infinity as x grows. As was discussed above, the empirical function can have either convex or concave form in its upper end. That is why three different functions were suggested for this second term: $a x$, $a x^2$ and $a \sqrt{x}$. The full expressions for the transformations then are:

$$u(x) = a x - b/x \quad (3.31)$$

$$u(x) = a x^2 - b/x \quad (3.32)$$

$$u(x) = a \sqrt{x} - b/x \quad (3.33)$$

In addition to the five two parameter transformations mentioned, the three parameter transformation suggested by Kartvelishvili (eq 3.22) was tested.

3.4.4 Transformations of the gamma distribution

The one parameter gamma distribution is not flexible enough to describe runoff phenomena. To improve the flexibility a second scale parameter can be introduced by replacing the variable x in eq (3.14) by the new variable x/β , which gives a two parameter gamma distribution function:

$$F(x; \beta, \gamma) = 1 / (\beta^\gamma \Gamma(\gamma)) \int_0^x t^{\gamma-1} \exp(-t/\beta) dt \quad (3.34)$$

The expected value, variation coefficient and skew coefficient of this distribution are $E(x) = \beta\gamma$, $C_v = 1/\sqrt{\gamma}$ and $C_s = 2/\sqrt{\gamma}$. As we are considering the module of flow the expected value is equal to one. The relation between the two parameters is thus $\beta = \gamma^{-1}$, which being inserted in the equation (3.34) gives the alternative form for the two parameter gamma distribution:

$$F(x; \gamma, \alpha_1) = \gamma^\gamma / \Gamma(\gamma) \int_0^x t^{\gamma-1} \exp(-\gamma t) dt \quad (3.35)$$

The parameters are γ and the first order moment of the origin α_1 . Kritskij and Menkel (1950) have suggested a three parameter gamma distribution by introducing the new variable x^b instead of x in this latter equation. We get the following formula:

$$F(x; \gamma, \alpha_1, b) = (\Gamma(\gamma + b) / \Gamma(\gamma))^{1/b} \int_0^x t^{\gamma/b-1} / (|b| \Gamma(\gamma)) \cdot \exp\{-\Gamma(\gamma + b) / \Gamma(\gamma) t\}^{1/b} dt \quad (3.36)$$

A three parameter gamma distribution can be also derived by the assumption of a lower bound ϵ for the variable x . If x in the eq (3.14) is replaced by the new variable $(x - \epsilon) / \beta$, then

$$F(x; \beta, \gamma, \epsilon) = 1 / (\beta^\gamma \Gamma(\gamma)) \int_0^x (t - \epsilon)^{\gamma-1} \exp(-(t - \epsilon) / \beta) dt \quad (3.37)$$

It can be shown (Yevjevich, 1972b) that this distribution can be written in the form of a Pearson type III distribution.

Empirical transformation functions of a gamma distribution cannot be constructed in the way it has been done for the normal distribution. The reason for this is that the gamma distribution does not have any form corresponding to the unit normal distribution.

3.4.5 Parameter estimation

Several methods are advised by statisticians for parameter estimation. Estimates found by different methods are classified with respect to their properties. A good estimate should have the following three properties, of which the first two are of theoretical and the third of practical nature (Aitchison and Brown 1957).

1. The estimate should be unbiased, or, when only large samples are in question, asymptotically unbiased.
2. The variance (or some similar measure) should be as small as possible (efficient).
3. The calculations involved should be reasonable and within the capabilities of the available computing machinery.

An estimate seldom answers all the desirable qualities. Estimates which are satisfactory with respect to the theoretical criteria tend to require much calculations. Of the methods available for estimation we distinguish five types.

1. the method of maximum likelihood
2. the method of moments
3. the method of quantiles
4. the graphical method
5. mixed methods

The oldest of these is the method of moments, introduced by K Pearson (Cramer, 1948). It often leads to comparatively simple calculations for determination of parameters. This can be one of the reasons for its wide application in hydrologic practice. In designing practice in the USSR it was the almost exclusively employed method (Kritskij and Menkel, 1964). Several other authors have, however, pointed out that estimates derived by this method can have considerably higher variance than estimates derived by the method of maximum likelihood (Aitchison and Brown (1957), Blochinov (1964), Yevjevich (1972b)). Further, the estimates can be highly biased (Wallis et al 1974). This first of all

accounts for high variation coefficients and skew coefficients. Blochinov does not advise therefore the method of moments for the value of the variation coefficient higher than 0,3, when estimating parameters to the gamma distributions eq 3.35 and eq 3.36. From examples given by Aitchison and Brown, almost the same conclusions can be drawn about the use of the method of moments to estimate parameters for the two and three parameter lognormal distributions (eq 3.20 and 3.21). Alekseev (1971) has suggested a method for corrections of systematical bias when using the method of moments for both gamma and lognormal distribution. Bias factors for different sample sizes for some common distributions including the lognormal and the gamma distribution have been published by Wallis et al (1974). To calculate these factors the Monte-Carlo Method has been applied. The same approach has been used by Reznikovskij (1969) for the Pearson type III distribution.

The graphical method with the help of probability papers is also widely applied in hydrology and especially - in flood frequency analysis. Parameter estimation was treated as a problem of curve fitting (Chow 1964) and for a long time the method of adapting a straight line to data, plotted on a probability paper (Dalrymple 1960) was considered to be the best one. The accuracy of this method is highly subjective, and depends on experience and good judgement. It is, however, advisable as a control measure, when other methods have been used. Gumbel (1958) uses the method of least squares to fit a straight line on a probability paper and thus eliminate the subjective aspect of the graphical method.

Hydrologic applications of the method of quantiles is shown, for instance, by Alekseev (1960). This method has advantageous properties. It is, first of all, very simple in its application. Further, it gives a variance of the estimates that usually does not differ very much from that, obtained by the maximum likelihood method (Aitchison and Brown (1957), Gumbel (1958)). The method can therefore represent an alternative to the maximum likelihood method when the amount of data is very large or the calculation process is very complicated. Johnson (1949a) recommends this method as the best one for his system of transformed normal distributions.

From the theoretical point of view, the most important general method of estimation is the method of maximum likelihood. The estimates can be biased but usually not considerably, which in most cases can be eliminated by suitable corrections. The maximum likelihood method was first applied in hydrology by Kritskij and Menkel (1949). The expressions for the determination of parameters are often complicated and of an implicit form, which leads to time consuming calculations. This is, first of all, valid for distribution functions with three or more parameters.

3.4.5.1 The method of maximum likelihood

For a sample $x^1, \dots, x^N$ of the random variable X characterized by its density function $f(\theta_1, \dots, \theta_k)$, where $\theta_1, \dots, \theta_k$ are the parameters, the sample distribution is given by (for the case of independence):

$$f_1, \dots, f_N(x^1, \dots, x^N) = f(x^1) f(x^2) \dots f(x^N)$$

This expression generally gives the probability of occurrence of different samples for given values of the parameters $\theta_1, \dots, \theta_k$. For our case the values of the parameters are unknown, while the sample is known. The most probable estimates of $\theta_1, \dots, \theta_k$ will be those maximizing the probability that the given sample should occur i.e. that will maximize the function:

$$L(\theta_1, \dots, \theta_k) = f(x^1; \theta_1, \dots, \theta_k) \dots f(x^N; \theta_1, \dots, \theta_k) \quad (3.38)$$

which is called the likelihood function. If the maximum can be located by differentiation, the maximum likelihood estimates are the solutions of the k equations:

$$\partial L(\theta_1, \dots, \theta_k) / \partial \theta_j = 0, \quad j = 1, \dots, k \quad (3.39)$$

Sometimes, it can be more convenient to maximize the log likelihood function $l(\theta_1, \dots, \theta_k) = \ln L(\theta_1, \dots, \theta_k)$. The maximum likelihood equations are then:

$$\partial \ln L / \partial \theta_j = \partial (\ln L) / \partial \theta_j = 1 / L \partial \ln L / \partial \theta_j, \quad j = 1, \dots, k \quad (3.40)$$

Equations for parameter estimation by the maximum likelihood method and comments on their solution for the normal distribution (eq 3.10) the two and three parameter lognormal (eq 3.20 and eq 3.21) and the two and three parameter gamma distributions (eq 3.34, eq 3.36 and eq 3.37) are given in Cohen (1951), Aitchison and Brown (1957), Thom (1958), Blochinov (1964) and Markovic (1965). These equations will be given below with only brief comments.

Normal distributions with parameters m and σ :

$$\hat{m} = \bar{x} = 1 / N \sum_{i=1}^N x^i \quad (3.41)$$

$$s = (1 / N \sum_{i=1}^N (x^i - \bar{x})^2)^{1/2} \quad (3.42)$$

Lognormal distribution with two parameters m_n and σ_n :

$$\hat{m}_n = \overline{\ln x} = 1 / N \sum_{i=1}^N \ln x^i \quad (3.43)$$

$$s_n = (1 / N \sum_{i=1}^N (\ln x^i - \hat{m}_n)^2)^{1/2} \quad (3.44)$$

Lognormal distribution with three parameters m_n , σ_n and ϵ :

$$\hat{m}_n = 1 / N \sum_{i=1}^N \ln (x^i - \epsilon) \quad (3.45)$$

$$s_n = (1 / N \sum_{i=1}^N (\ln (x^i - \epsilon) - \hat{m}_n)^2)^{1/2} \quad (3.46)$$

$$\begin{aligned} & \sum_{i=1}^N (x^i - \epsilon)^{-1} \cdot \{1 / N \sum_{i=1}^N \ln^2 (x^i - \epsilon) - \\ & (1 / N \sum_{i=1}^N \ln (x^i - \epsilon))^2 - 1 / N \sum_{i=1}^N \ln (x^i - \epsilon)\} + \\ & + \sum_{i=1}^N \{\ln (x^i - \epsilon) / (x^i - \epsilon)\} = 0 \end{aligned} \quad (3.47)$$

ϵ have to be obtained from eq (3.47) with an iterative procedure first and then $\hat{m}_n$ and s_n can be obtained from equations (3.45) and (3.46).

Gamma distribution with two parameters β and γ : The estimate of the parameter γ is given of the equation

$$\partial (\ln \Gamma(\hat{\gamma})) / \partial \hat{\gamma} - \ln \hat{\gamma} - \lambda = 0 \quad (3.48)$$

where

$$\lambda = \sum_{i=1}^N \ln(x_i / \bar{x}) / N$$

Kritskij and Menkel (1949) have tabulated the relation between $\hat{\gamma}$ and λ . Thom (1958) gives an explicit expression for $\hat{\gamma}$ with the help of asymptotic expansion for the digamma function $\partial (\ln \Gamma(\hat{\gamma})) / \partial \hat{\gamma}$. The following expression is derived:

$$\hat{\gamma} = \{1 + \{1 + 4/3 (\ln \bar{x} - 1 / N \sum_{i=1}^N \ln x_i)\}^{1/2}\} / \{4 (\ln \bar{x} - 1 / N \sum_{i=1}^N \ln x_i)\} - \Delta \hat{\gamma} \quad (3.49)$$

where $\Delta \hat{\gamma}$ is a correction factor depending on the fact, that only one term in the digamma expansion was considered. A table for $\Delta \hat{\gamma}$ is given by Thom. The parameter β is calculated from the expression

$$\hat{\beta} = \bar{x} / \hat{\gamma} \quad (3.50)$$

As we are using the modular coefficients ($x_i = Q_i / \bar{Q}$), the average $\bar{x}$ is equal to one, which was the condition used, to derive the distribution eq (3.35).

Gamma distribution with three parameters β , γ and ϵ :

The parameter γ and β can be estimated from expressions in accordance with eqs. 3.49 and 3.50

$$\hat{\gamma} = \{1 + \{1 + 4/3 (\ln (\bar{x} - \hat{\epsilon}) - 1 / N \sum_{i=1}^N \ln (x_i - \hat{\epsilon}))\}^{1/2}\} / \{4 (\ln (\bar{x} - \hat{\epsilon}) - 1 / N \sum_{i=1}^N \ln (x_i - \hat{\epsilon}))\} \quad (3.51)$$

$$\hat{\beta} = (\bar{x} - \hat{\epsilon}) / \hat{\gamma} \quad (3.52)$$

Estimate of the lower bound $\hat{\epsilon}$ is obtained by an iterative procedure from the implicit expression:

$$\begin{aligned} & (1 + \{1 + 4/3 (\ln (\bar{x} - \hat{\epsilon}) - 1 / N \sum_{i=1}^N (x^i - \hat{\epsilon}))\}^{1/2}) \\ & (1 + \{1 + 4/3 (\ln (\bar{x} - \hat{\epsilon}) - 1 / N \sum_{i=1}^N \ln (x^i - \hat{\epsilon}))\}^{1/2} - \\ & - 4 \ln (\bar{x} - \hat{\epsilon}) - 1 / N \sum_{i=1}^N \ln (x^i - \hat{\epsilon})) - \\ & (\bar{x} - \hat{\epsilon}) 1 / N \sum_{i=1}^N (x^i - \hat{\epsilon})^{-1} = 0 \end{aligned} \quad (3.53)$$

Gamma distribution with three parameters α_1 , γ and b :

The likelihood equations are

$$\begin{aligned} & \{ \Gamma (\hat{\gamma} + \hat{b}) / \Gamma (\hat{\gamma}) \}^{1/\hat{b}} \lambda_1 - \hat{\gamma} = 0 \\ & \lambda_2 + \ln \{ \Gamma (\hat{\gamma} + \hat{b}) / \Gamma (\hat{\gamma}) \} - \hat{b} \partial (\ln \Gamma (\hat{\gamma})) / \partial \hat{\gamma} = 0 \\ & \hat{\gamma} \lambda_2 - \{ \Gamma (\hat{\gamma} + \hat{b}) / \Gamma (\hat{\gamma}) \}^{1/\hat{b}} \lambda_3 + \hat{b} = 0 \end{aligned}$$

where

$$\begin{aligned} \lambda_1 &= 1 / N \sum_{i=1}^N (x^i / \hat{\alpha}_1)^{1/\hat{b}} \\ \lambda_2 &= 1 / N \sum_{i=1}^N \ln (x^i / \hat{\alpha}_1) \\ \lambda_3 &= 1 / N \sum_{i=1}^N (x^i / \hat{\alpha}_1) \ln (x^i / \hat{\alpha}_1) \end{aligned}$$

A solution of the transcendental system (3.54) involves technical difficulties, even if a computer is employed. Blochinov (1964) suggests an approximate method, based on two simplifications:

1. the arithmetical mean $\bar{x}$ which has high effectiveness within a wide range of C_v and $\hat{C}_s$ is assumed to be an estimate for α_1
2. likelihood estimates of the other two parameters are found by dissecting the likelihood surface into planes and then maximum is found for each plane. The greatest value of the likelihood

function is chosen out of these maxima and solution is obtained.

For further details see, for instance, Kartvelishvili (1967).

We shall now turn to the transformations of the unit normal distribution given in part 3.4.3. All transformation functions can be written in the form:

$$u(x) = \sum_{j=1}^k \alpha_j \phi_j(x) \quad (3.55)$$

where k is two and three, respectively. The density distribution function is, in accordance with eq 3.16, given by:

$$f(x) = 1 / \sqrt{2\pi} \exp \left\{ -\frac{1}{2} \left\{ \sum_{j=1}^k \alpha_j \phi_j(x) \right\}^2 \right\} \sum_{j=1}^k \alpha_j \phi_j'(x) \quad (3.56)$$

The log likelihood function is:

$$\begin{aligned} l(\alpha_1, \dots, \alpha_k) = & -N \ln \sqrt{2\pi} - 1/2 \sum_{i=1}^N \left(\sum_{j=1}^k \alpha_j \phi_j(x^i) \right)^2 + \\ & + \sum_{j=1}^k \ln \sum_{j=1}^k \alpha_j \phi_j'(x^i) \end{aligned}$$

Differentiating eq (3.57) in respect to the parameters α_m , $m = 1, \dots, k$ we derive the k likelihood equations:

$$\begin{aligned} - \sum_{i=1}^N \left\{ \sum_{j=1}^k \hat{\alpha}_j \phi_j(x^i) \right\} \phi_m'(x^i) + \sum_{i=1}^N \left\{ \sum_{j=1}^k \hat{\alpha}_j \phi_j'(x^i) \right\}^{-1} \phi_m'(x^i) = 0 \\ , m = 1, \dots, k \end{aligned} \quad (3.58)$$

from which the estimates α_m , $m = 1, \dots, k$ can be calculated.

Let us now insert the special expression for the transformation functions in the likelihood equations (3.58). Explicit expression for the estimated parameters cannot be found in all cases, the likelihood equation will be then only rearranged to a convenient form.

$$u(x) = a x - b:$$

$$\hat{a} = (1/N \{ \sum_{i=1}^N (x^i)^2 - 1/N (\sum_{i=1}^N x^i)^2 \})^{-1/2} \quad (3.59)$$

$$\hat{b} = \hat{a} / N \sum_{i=1}^N x^i \quad (3.60)$$

$$u(x) = a \ln x - b:$$

$$\hat{a} = (1/N \{ \sum_{i=1}^N \ln^2 x^i - 1/N (\sum_{i=1}^N \ln x^i)^2 \})^{-1/2} \quad (3.61)$$

$$\hat{b} = \hat{a} / N \sum_{i=1}^N \ln x^i \quad (3.62)$$

$$u(x) = a x - b/x:$$

$$\begin{aligned} -\hat{a} \sum_{i=1}^N (x^i)^2 + N \cdot \hat{b} + \sum_{i=1}^N (\hat{a} + \hat{b}/(x^i)^2)^{-1} &= 0 \\ -N \hat{a} + \hat{b} \sum_{i=1}^N (x^i)^{-2} + \sum_{i=1}^N (\hat{a} (x^i)^2 + \hat{b})^{-1} &= 0 \end{aligned} \quad (3.63)$$

$$u(x) = a x^2 - b/x:$$

$$\begin{aligned} -\hat{a} \sum_{i=1}^N (x^i)^4 + \hat{b} \sum_{i=1}^N x^i + \sum_{i=1}^N (\hat{a} + \hat{b}/2(x^i)^3)^{-1} &= 0 \\ -\hat{a} \sum_{i=1}^N x^i + \hat{b} \sum_{i=1}^N (x_i)^{-2} + \sum_{i=1}^N (2\hat{a}(x_i)^3 + \hat{b})^{-1} &= 0 \end{aligned} \quad (3.64)$$

$$u(x) = a \sqrt{x} - b/x:$$

$$\begin{aligned} -\hat{a} \sum_{i=1}^N x^i + \hat{b} \sum_{i=1}^N (x^i)^{-1/2} + \sum_{i=1}^N (\hat{a} + 2\hat{b}(x^i)^{-3/2})^{-1} &= 0 \\ -\hat{a} \sum_{i=1}^N (x^i)^{-1/2} + \hat{b} \sum_{i=1}^N (x^i)^{-2} + \sum_{i=1}^N (\hat{a}(x^i)^{3/2} / 2 + \hat{b})^{-1} &= 0 \end{aligned} \quad (3.65)$$

$$u(x) = a \ln(\ln(1+x)) + b \ln(x) + c \sqrt{x};$$

$$\begin{aligned} & - \sum_{i=1}^N A^i \ln(\ln(1+x^i)) + \sum_{i=1}^N B^i / \{(1+x^i) \ln(1+x^i)\} = 0 \\ & - \sum_{i=1}^N A^i \ln(x^i) + \sum_{i=1}^N B^i / x^i = 0 \\ & - \sum_{i=1}^N A^i \sqrt{x^i} + \sum_{i=1}^N B^i / 2 \sqrt{x^i} = 0 \end{aligned} \quad (3.66)$$

where

$$\begin{aligned} A^i &= a \ln(\ln(1+x^i)) + b \ln(x^i) + c \sqrt{x^i} \\ B^i &= a / \{\ln(1+x^i) (1+x^i)\} + b / x^i + c / (2 \sqrt{x^i}) \end{aligned}$$

For the two first transformation functions, explicit equations for the estimates with the help of the likelihood method are obtained and these can be thus easily calculated. In all the other cases the estimates have to be computed by some iterative method. An alternative method is to search for the maximum of the likelihood equation directly. This latter method was considered to be the most suitable and was accepted for this study. Numerically the problem of finding the maximum of the log likelihood function was solved by applying the method of steepest descent and simple scanning. The topography of the log likelihood function was in some cases very flat in the area of interest. This is of special significance for the three parameter curve, most of all in respect to the first parameter a .

3.4.5.2 The method of moments

The method of moments consists in equating a convenient number of the empirical moments to the corresponding moments of the theoretical distribution, which are functions of the unknown parameters. The number of empirical moments, which must be computed, is equal to the number of unknown parameters.

For the theoretical moments of the normal, lognormal and gamma distributions we refer to Aitchison and Brown (1957), Kartvelishvili (1967) and Yevjevich (1970b). This method was not used for the calculation of parameters of these distributions and will not be commented further.

Application of the method of moments of transformations of the unit normal distribution to the form expressed by eq (3.55) can lead to mathematical and numerical difficulties, as the theoretical moments cannot be given with a simple expression. To avoid this problem we suggest a modification of the method of moments.

In order to do this we set the theoretical moments of the unit normal distribution equal to the moments of the transformed sample. The system of equations for the case of a two parameter transformation function is:

$$1 / N \sum_{i=1}^N u(x^i) = 0 \quad (3.67)$$

$$1 / N \sum_{i=1}^N (u(x^i))^2 = 1 \quad (3.68)$$

When we have a three parameter transformation a third equation derived from the third moment should be added:

$$1 / N \sum_{i=1}^N (u(x^i))^3 = 0 \quad (3.69)$$

Inserting the expression eq (3.55) in equations (3.67) and (3.68) and opening the brackets, we get for the two parameter transformation:

$$\hat{\alpha}_1 / N \sum_{i=1}^N \phi_1(x^i) + \hat{\alpha}_2 / N \sum_{i=1}^N \phi_2(x^i) = 0 \quad (3.70)$$

$$\begin{aligned} \hat{\alpha}_1^2 / N \sum_{i=1}^N \phi_1^2(x^i) + 2 \hat{\alpha}_1 \hat{\alpha}_2 / N \sum_{i=1}^N \phi_1(x^i) \phi_2(x^i) + \\ + \hat{\alpha}_2^2 / N \sum_{i=1}^N \phi_2^2(x^i) = 1 \end{aligned} \quad (3.71)$$

Solving the first equation for $\hat{\alpha}_2$ we find:

$$\hat{\alpha}_2 = - \hat{\alpha}_1 \sum_{i=1}^N \phi_1(x^i) / \sum_{i=1}^N \phi_2(x^i) \quad (3.72)$$

which when inserted in the second equation (3.71) will give the following second order equation for $\hat{\alpha}_1$.

$$\hat{\alpha}_1 \sum_{i=1}^N \phi_1^2(x^i) - 2 \hat{\alpha}_1^2 \sum_{i=1}^N \phi_1(x^i) \phi_2(x^i) \cdot \sum_{i=1}^N \phi_1(x^i) / \phi_2(x^i) + \hat{\alpha}_1^2 \{ \sum_{i=1}^N \phi_1(x^i) / \sum_{i=1}^N \phi_2(x^i) \} \sum_{i=1}^N \phi_2^2(x^i) = N \quad (3.73)$$

This equation has the positive root:

$$\hat{\alpha}_1 = \{ \{ \sum_{i=1}^N \phi_1^2(x^i) - 2 \sum_{i=1}^N \phi_1(x^i) \phi_2(x^i) \sum_{i=1}^N \phi_1(x^i) / \sum_{i=1}^N \phi_2(x^i) + (\sum_{i=1}^N \phi_1(x^i) / \sum_{i=1}^N \phi_2(x_i))^2 \sum_{i=1}^N \phi_2^2(x_i) / N \}^{-1/2} \} \quad (3.74)$$

The estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are thus calculated from equations (3.74) and (3.72). Inserting the specific expressions for the transformation functions eq (3.29 - 3.33) we find:

$$u(x) = a \cdot x - b:$$

$$\hat{a} = \{ \{ (\sum_{i=1}^N (x^i)^2 - (\sum_{i=1}^N x^i)^2 / N) / N \}^{-1/2} = 1/s \quad (3.59)$$

$$\hat{b} = \hat{a} \sum_{i=1}^N x^i / N = \hat{a} \bar{x} \quad (3.60)$$

$$u(x) = a \cdot \ln(x) - b:$$

$$\hat{a} = \{ \{ (\sum_{i=1}^N \ln^2 x^i - (\sum_{i=1}^N \ln x^i)^2 / N) / N \}^{-1/2} = 1/s_n \quad (3.61)$$

$$\hat{b} = \hat{a} \sum_{i=1}^N \ln x^i / N = \hat{a} \overline{\ln x} \quad (3.62)$$

$$u(x) = a \cdot x - b / x:$$

$$\hat{a} = \{ \{ (\sum_{i=1}^N (x^i)^2 - 2N \sum_{i=1}^N x^i / \sum_{i=1}^N 1/x^i + (\sum_{i=1}^N x^i / \sum_{i=1}^N 1/x^i)^2 \sum_{i=1}^N (x^i)^{-2} / N \}^{-1/2} \} \quad (3.75)$$

$$\hat{b} = \hat{a} \sum_{i=1}^N 1 / x^i \quad (3.76)$$

$$u(x) = a x^2 - b / x:$$

$$\begin{aligned} \hat{a} = \{ & \left(\sum_{i=1}^N (x^i)^4 - 2 \sum_{i=1}^N x^i \sum_{i=1}^N (x^i)^2 / \sum_{i=1}^N 1 / x^i + \right. \\ & \left. \left(\sum_{i=1}^N (x^i)^2 / \sum_{i=1}^N 1 / x^i \right)^2 \sum_{i=1}^N (x^i)^{-2} \right) / N \}^{-1/2} \end{aligned} \quad (3.77)$$

$$\hat{b} = \hat{a} \sum_{i=1}^N (x^i)^2 / \sum_{i=1}^N 1/x^i \quad (3.78)$$

$$u(x) = a \sqrt{x} - b / x:$$

$$\begin{aligned} \hat{a} = \{ & \left(\sum_{i=1}^N x^i - 2 \sum_{i=1}^N (x^i)^{-1/2} \sum_{i=1}^N (x^i)^{1/2} / \sum_{i=1}^N 1 / x^i + \right. \\ & \left. \left(\sum_{i=1}^N (x^i)^{1/2} / \sum_{i=1}^N 1 / x^i \right)^2 \sum_{i=1}^N 1 / (x^i)^2 \right) / N \}^{-1/2} \end{aligned} \quad (3.79)$$

$$\hat{b} = \hat{a} \sum_{i=1}^N (x^i)^{1/2} / \sum_{i=1}^N 1 / x^i \quad (3.80)$$

It should be noted, that for the two first transformations by this modified moment method, the estimates are the same as those got by the conventional method of moments and also by the method of maximum likelihood. For the three other transformations the calculation of parameter estimates is relatively simple, compared to the method of maximum likelihood, as in this case explicit expressions for the parameters are found.

For a three parameter transformation of the unit normal distribution we can write down in accordance with eqs (3.67-69) the following three equations for the derivations of the parameter estimates:

$$\hat{\alpha}_1 \sum_{i=1}^N \phi_1(x^i) + \hat{\alpha}_2 \sum_{i=1}^N \phi_2(x^i) + \hat{\alpha}_3 \sum_{i=1}^N \phi_3(x^i) = 0 \quad (3.81)$$

$$\begin{aligned}
& \hat{\alpha}_1^2 \sum_{i=1}^N \phi_1^2(x^i) + \hat{\alpha}_2^2 \sum_{i=1}^N \phi_2^2(x^i) + \hat{\alpha}_3^2 \sum_{i=1}^N \phi_3^2(x^i) + \\
& 2 \hat{\alpha}_1 \hat{\alpha}_2 \sum_{i=1}^N \phi_1(x^i) \phi_2(x^i) + 2 \hat{\alpha}_1 \hat{\alpha}_3 \sum_{i=1}^N \phi_1(x^i) \phi_3(x^i) + \\
& 2 \hat{\alpha}_2 \hat{\alpha}_3 \sum_{i=1}^N \phi_2(x^i) \phi_3(x^i) = N
\end{aligned} \quad (3.82)$$

$$\begin{aligned}
& \hat{\alpha}_1^3 \sum_{i=1}^N \phi_1^3(x^i) + \hat{\alpha}_2^3 \sum_{i=1}^N \phi_2^3(x^i) + \hat{\alpha}_3^3 \sum_{i=1}^N \phi_3^3(x^i) + \\
& 3 \hat{\alpha}_1^2 \hat{\alpha}_2 \sum_{i=1}^N \phi_1^2(x^i) \phi_2(x^i) + 3 \hat{\alpha}_1^2 \hat{\alpha}_3 \sum_{i=1}^N \phi_1^2(x^i) \phi_3(x^i) + \\
& 3 \hat{\alpha}_1 \hat{\alpha}_2^2 \sum_{i=1}^N \phi_1(x^i) \phi_2^2(x^i) + 3 \hat{\alpha}_2^2 \hat{\alpha}_3 \sum_{i=1}^N \phi_2^2(x^i) \phi_3(x^i) + \\
& 3 \hat{\alpha}_1 \hat{\alpha}_3^2 \sum_{i=1}^N \phi_1(x^i) \phi_3^2(x^i) + 3 \hat{\alpha}_2 \hat{\alpha}_3^2 \sum_{i=1}^N \phi_2(x^i) \phi_3^2(x^i) + \\
& 6 \hat{\alpha}_1 \hat{\alpha}_2 \hat{\alpha}_3 \sum_{i=1}^N \phi_1(x^i) \phi_2(x^i) \phi_3(x^i) = 0
\end{aligned} \quad (3.83)$$

From equation (3.81) we solve $\hat{\alpha}_3$ i.e.

$$\hat{\alpha}_3 = - \hat{\alpha}_1 \frac{\sum_{i=1}^N \phi_1(x^i)}{\sum_{i=1}^N \phi_3(x^i)} - \hat{\alpha}_2 \frac{\sum_{i=1}^N \phi_2(x^i)}{\sum_{i=1}^N \phi_3(x^i)} \quad (3.84)$$

Inserting this expression in the two other equations we can find after some calculations the following equation system of α_1 and α_2 :

$$\sum_{k=1}^2 \sum_{m=1}^2 A_{km} \hat{\alpha}_k \hat{\alpha}_m - N = 0 \quad (3.85)$$

$$\sum_{j=1}^2 \sum_{k=1}^2 \sum_{m=1}^2 A_{jkm} \hat{\alpha}_j \hat{\alpha}_k \hat{\alpha}_m = 0$$

where

$$A_{km} = \sum_{i=1}^N \phi_k(x^i) \phi_m(x^i) - \left[\sum_{i=1}^N \phi_k(x^i) / \sum_{i=1}^N \phi_3(x^i) \right] \sum_{i=1}^N \phi_m(x^i) \phi_3(x^i) - \left[\sum_{i=1}^N \phi_m(x^i) / \sum_{i=1}^N \phi_3(x^i) \right] \sum_{i=1}^N \phi_k(x^i) \phi_3(x^i) + \left[\sum_{i=1}^N \phi_k(x^i) \sum_{i=1}^N \phi_m(x^i) / \left(\sum_{i=1}^N \phi_3(x^i) \right)^2 \right] \sum_{i=1}^N \phi_3^2(x^i)$$

$$k = 1, 2 \quad m = 1, 2$$

$$A_{jkm} = \sum_{i=1}^N \phi_j(x^i) \phi_k(x^i) \phi_m(x^i) - \left[\sum_{i=1}^N \phi_j(x^i) \sum_{i=1}^N \phi_k(x^i) \sum_{i=1}^N \phi_m(x^i) / \left(\sum_{i=1}^N \phi_3(x^i) \right)^3 \right] \sum_{i=1}^N \phi_3^3(x^i) + \left[\sum_{i=1}^N \phi_j(x^i) \sum_{i=1}^N \phi_k(x^i) / \left(\sum_{i=1}^N \phi_3(x^i) \right)^2 \right] \sum_{i=1}^N \phi_m(x^i) \phi_3(x^i)^2 + \left[\sum_{i=1}^N \phi_m(x^i) \sum_{i=1}^N \phi_j(x^i) / \left(\sum_{i=1}^N \phi_3(x^i) \right)^2 \right] \sum_{i=1}^N \phi_k(x^i) \phi_3(x^i)^2 + \left[\sum_{i=1}^N \phi_k(x^i) \sum_{i=1}^N \phi_m(x^i) / \left(\sum_{i=1}^N \phi_3(x^i) \right)^2 \right] \sum_{i=1}^N \phi_j(x^i) \phi_3(x^i)^2 - \left[\sum_{i=1}^N \phi_j(x^i) / \sum_{i=1}^N \phi_3(x^i) \right] \sum_{i=1}^N \phi_k(x^i) \phi_m(x^i) \phi_3(x^i) - \left[\sum_{i=1}^N \phi_m(x^i) / \sum_{i=1}^N \phi_3(x^i) \right] \sum_{i=1}^N \phi_j(x^i) \phi_k(x^i) \phi_3(x^i) - \left[\sum_{i=1}^N \phi_k(x^i) / \sum_{i=1}^N \phi_3(x^i) \right] \sum_{i=1}^N \phi_m(x^i) \phi_j(x^i) \phi_3(x^i)$$

By inserting the expression for the three parameter transformation eq (3.22) in the equation system above (3.85), the parameters a , b and c can be obtained. Numerically this can be done by a method described in Fröberg (1962). From the two functions:

$$G_1(\hat{\alpha}_1, \hat{\alpha}_2) = \sum_{k=1}^2 \sum_{m=1}^2 A_{km} \hat{\alpha}_k \hat{\alpha}_m - N$$

$$G_2(\hat{\alpha}_1, \hat{\alpha}_2) = \sum_{j=1}^2 \sum_{k=1}^2 \sum_{m=1}^2 A_{jkm} \hat{\alpha}_j \hat{\alpha}_k \hat{\alpha}_m$$

where now $\hat{\alpha}_1$ and $\hat{\alpha}_2$ stands for $\hat{a}$ and $\hat{b}$, respectively, we write down the new function:

$$g = G_1^2 + G_2^2$$

A solution to the equation system (3.85) will correspond to a minimum of the function g with the value zero. The problem of solving the equation system is, thus, reversed to that of finding minimum of a two dimensional function, which can be done by conventional methods. The equation system has at least two solutions. We can see this from an example, calculated from the series 53-121 Fäggeby for April. The following two equations are derived:

$$0.0202 \cdot \hat{a}^3 + 0.0208 \hat{b}^3 + 0.0621 \hat{a} \hat{b}^2 + 0.0615 \hat{a}^2 \hat{b} = 0$$

$$0.1381 \hat{a}^2 + 0.1693 \hat{b}^2 + 0.3012 \hat{a} \hat{b} = 1$$

A solution is $a = -10.395$ and $b = 11.990$ and we can at once see that $a = 10.395$ and $b = -11.990$ is also a solution. Only the first solution is, however, applicable. It is apparent that the computations are not simplified greatly, in comparison with maximum likelihood method, by applying the modified moment method in order to find estimates for the three parameter transformation.

3.4.5.3 The graphical method

The graphical method consists in visual fitting of a curve to a plotted empirical distribution on a graph and selecting the number of points from this curve in coincidence with the number of parameters, from which estimates of the parameters can be calculated, usually a transformation of the scale is performed in such way, that the fitted curve will form a straight line.

We shall modify this method so, that the curve fitting can be done by the method of least squares. Gumbel (1958) gives an example of how to calculate estimates for the linear transformation eq (3.29) of the unit normal distribution in this way. From part 3.4.2 we know that the empirical transformation curve can be represented by a set of points (x^i, u^i) , $i = 1, \dots, N$. Fitting a straight line to these points by the method of least squares and minimizing the horizontal distances lead to the following equations for estimates of the parameters $\hat{a}'$ and $\hat{b}'$:

$$\hat{a}' = \left(\sum_{i=1}^N u^i x^i - 1/N \sum_{i=1}^N u^i \sum_{i=1}^N x^i \right) / \left(\sum_{i=1}^N x^i - 1/N \sum_{i=1}^N x^i \right) = \left(\sum_{i=1}^N u^i x^i - N \bar{u} \bar{x} \right) / s_x^2$$

$$\hat{b}' = \left(\sum_{i=1}^N u^i + \hat{a}' \sum_{i=1}^N x^i \right) / N = \bar{u} + \hat{a}' \bar{x}$$

If we minimize instead the vertical distances, i.e. starting from the equation $x = (u - b) / a$, we find the estimates:

$$\hat{a}'' = \left(\sum_{i=1}^N (u^i)^2 - 1/N \left(\sum_{i=1}^N u^i \right)^2 \right) / \left(\sum_{i=1}^N u^i x^i - 1/N \sum_{i=1}^N u^i \sum_{i=1}^N x^i \right) = s_u^2 / \left(\sum_{i=1}^N u^i x^i - N \bar{u} \bar{x} \right)$$

$$\hat{b}'' = \left(\sum_{i=1}^N u^i + \hat{a}'' \sum_{i=1}^N x^i \right) / N = \bar{u} + \hat{a}'' \bar{x}$$

To eliminate the calculations of crossproducts Gumbel suggests new estimates as the geometrical means of the two estimates derived by minimizing of the horisontal and vertical distances, respectively i.e.

$$a = \sqrt{a' a''} \quad b - \bar{u} = \sqrt{(b' - \bar{u})(b'' - \bar{u})}$$

This gives the estimates:

$$\hat{a} = s_u / s_x \quad (3.86)$$

$$\hat{b} = \bar{u} + \hat{a} \bar{x} \quad (3.87)$$

The estimates are closely related to the method of moments, which can be seen from comparing equations (3.59) and (3.60) to the two equations above. They differ only by the parameters $\bar{u}$ and s_u , which in general case are dependent on the sample size N and the underlying distribution. They can thus be treated as corrections of bias due to finite sample size. Knowing the underlying distribution, $\bar{u}$ and s_u can be once and for all calculated for different N . For the normal distribution s_u has been calculated as a function of N with the help of the relation $u^i = \Phi^{-1}(p_i)$ and is represented in Fig 3.26. Three

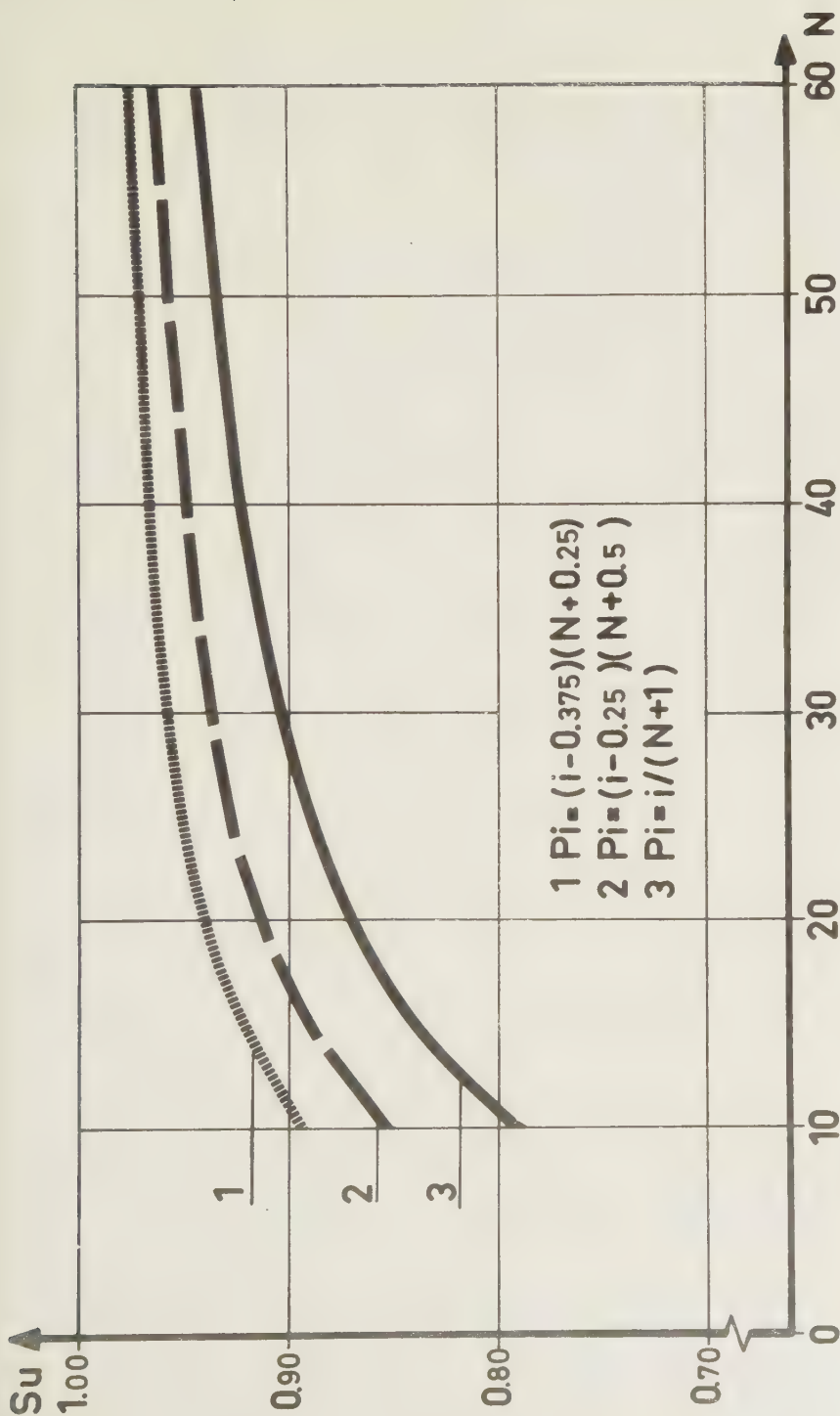


Fig 3.26 Standard deviation of transformed β -values S_u as a function of the number of observations N .

different formulas for p_i have been used, namely eqs (3.5), (3.8) and (3.9). Further on, we shall use eq (3.8). $\bar{u}$ is equal to zero for all N for the normal distribution.

For the transformation eq (3.30) we can similarly find the estimates:

$$\hat{a} = s_u / s_{\ln x} \quad (3.88)$$

$$\hat{b} = \hat{a} \cdot \overline{\ln x} \quad (3.89)$$

By minimizing the horizontal distances for the transformations eqs (3.31), (3.32), (3.33) and (3.22) the following equations for parameter estimation are got:

$$u(x) = ax - b / x:$$

$$\begin{aligned} \hat{a} = & \left(\sum_{i=1}^N u^i x^i - N \sum_{i=1}^N (u^i / x^i) / \sum_{i=1}^N 1 / x^i \right) / \left(\sum_{i=1}^N (x^i)^2 - \right. \\ & \left. - N^2 / \sum_{i=1}^N (x^i)^{-2} \right) \end{aligned} \quad (3.90)$$

$$\hat{b} = \left(\hat{a} \cdot N - \sum_{i=1}^N u^i / x^i \right) / \sum_{i=1}^N (x^i)^{-2} \quad (3.91)$$

$$u(x) = a x^2 - b / x:$$

$$\begin{aligned} \hat{a} = & \left(\sum_{i=1}^N u^i (x^i)^2 - \left[\left(\sum_{i=1}^N u^i / x^i \right) / \sum_{i=1}^N 1/x^i \right] \sum_{i=1}^N x^i \right) / \\ & \left(\sum_{i=1}^N (x^i)^4 - \left(\sum_{i=1}^N x^i \right)^2 / \sum_{i=1}^N (x^i)^{-2} \right) \end{aligned} \quad (3.92)$$

$$\hat{b} = \left(\hat{a} \sum_{i=1}^N x^i - \sum_{i=1}^N u^i / x^i \right) / \sum_{i=1}^N (x^i)^{-2} \quad (3.93)$$

$$u(x) = a \sqrt{x} - b/x:$$

$$\hat{a} = \left(\sum_{i=1}^N u^i \sqrt{x^i} - \left[\left(\sum_{i=1}^N u^i / x^i \right) / \sum_{i=1}^N (x^i)^{-2} \sum_{i=1}^N 1 / \sqrt{x^i} \right] \right) / \sum_{i=1}^N (x^i)^{-2} \quad (3.94)$$

$$\hat{b} = \left(\hat{a} \sum_{i=1}^N 1 / \sqrt{x^i} - \sum_{i=1}^N u^i / x^i \right) / \sum_{i=1}^N (x^i)^{-2} \quad (3.95)$$

$$u(x) = a \ln(\ln(1+x)) + b \ln(x) + c \sqrt{x}:$$

$$\hat{a} = \begin{vmatrix} A_{10} & A_{12} & A_{13} \\ A_{20} & A_{22} & A_{23} \\ A_{30} & A_{32} & A_{33} \end{vmatrix} \cdot B^{-1} \quad (3.96)$$

$$\hat{b} = \begin{vmatrix} A_{11} & A_{10} & A_{13} \\ A_{21} & A_{20} & A_{23} \\ A_{31} & A_{30} & A_{33} \end{vmatrix} \cdot B^{-1} \quad (3.97)$$

$$\hat{c} = \begin{vmatrix} A_{11} & A_{12} & A_{10} \\ A_{21} & A_{22} & A_{20} \\ A_{31} & A_{32} & A_{30} \end{vmatrix} \cdot B^{-1} \quad (3.98)$$

where

$$B = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}$$

$$A_{10} = \sum_{i=1}^N u^i \ln(\ln(1+x^i)) \quad A_{20} = \sum_{i=1}^N u^i \ln(x^i)$$

$$A_{30} = \sum_{i=1}^N u^i \sqrt{x^i}$$

$$A_{11} = \sum_{i=1}^N (\ln(\ln(1+x^i)))^2 \quad A_{12} = A_{21} = \sum_{i=1}^N \ln(\ln(1+x^i))$$

$$\ln(x^i)$$

$$\begin{aligned}
 A_{13} = A_{31} &= \sum_{i=1}^N \ln(\ln(1 + x^i)) \sqrt{x^i} & A_{22} &= \sum_{i=1}^N (\ln x^i)^2 \\
 A_{23} = A_{32} &= \sum_{i=1}^N \ln(x^i) \sqrt{x^i} & A_{33} &= \sum_{i=1}^N x^i
 \end{aligned}$$

The graphical method, where the method of least squares is used for curve fitting, thus gives relatively simple expressions for the parameter estimates. Further, the method involves a correction of bias due to finite sample size.

3.4.5.4 Conclusions

From the theoretical point of view the method of maximum likelihood will give the best estimates. The largest differences between estimates, calculated with this method and the graphical and moment methods, can be expected, when the variation and skew coefficients are high. An example of parameter estimates for six different transformation functions, calculated by the three different methods described above, are shown in tables 3.28 - 3.33. The 68 years long observation series from Dalälven at Fäggeby (53-121) has been used as an example. The discrepancy between the parameter values is relatively small for the two parameter transformations. Estimates found by the graphical method differs the most, due to the correction for bias. From the two examples drawn in Fig 3.27 for June we can see that the transformation curve calculated by this method always lies below the other two curves in the upper part of the graph and above in the lower part. For the other transformations the differences are the same as in the examples shown or smaller. The three parameter transformation eq (3.22) gives rise to large differences in the calculated parameter values. The parameters got by the graphical and the moment methods lie closer to each other than to the maximum likelihood parameters. It is also clear that the maximum likelihood parameters have stable positive values, while the other methods give large positive and negative values which compensate each other. The resulting curves can be expected to be less accurate. At the same time, from the example shown in Fig 3.27, we can see that the differences are not great. For other months they can be somewhat greater.

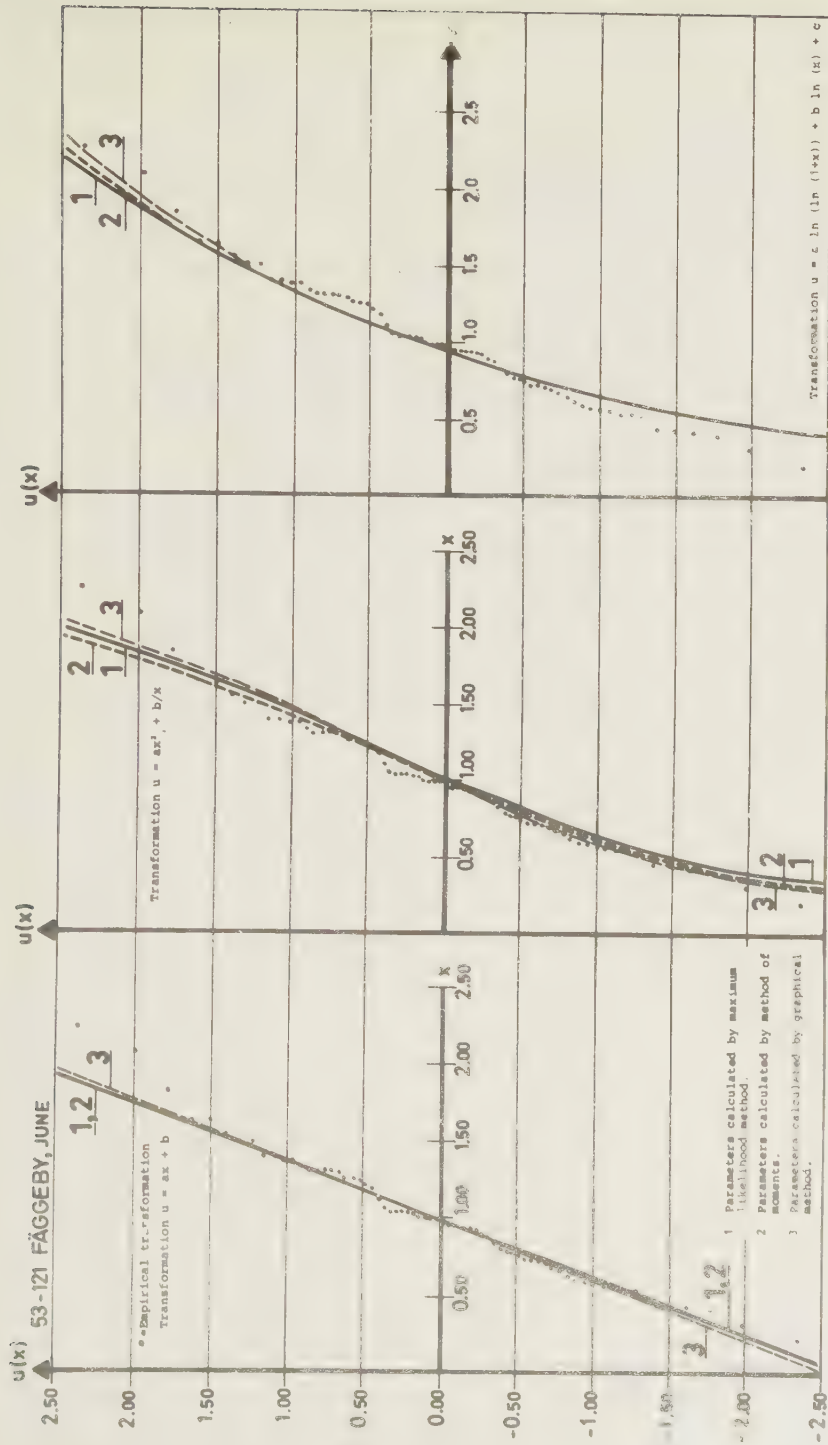


Fig 3.27 Examples of transformation curves of the unit normal distribution with parameters estimated with different methods.

TABLE 3.28 Estimated parameters for transformation:

$u(x) = ax - b$ with maximum likelihood, moment
and graphical methods

Month	Method					
	Maximum likelihood		Moment		Graphical	
	a	b	a	b	a	b
Jan	1.00	0.40	1.00	0.40	1.00	0.45
Feb	1.00	0.32	1.00	0.32	1.00	0.36
March	1.00	0.28	1.00	0.28	1.00	0.30
April	1.00	0.39	1.00	0.39	1.00	0.40
May	1.00	0.37	1.00	0.37	1.00	0.38
June	1.00	0.37	1.00	0.37	1.00	0.39
July	1.00	0.35	1.00	0.35	1.00	0.36
August	1.00	0.47	1.00	0.47	1.00	0.49
Sept	1.00	0.50	1.00	0.50	1.00	0.53
Oct	1.00	0.47	1.00	0.47	1.00	0.50
Nov	1.00	0.41	1.00	0.41	1.00	0.42
Dec	1.00	0.35	1.00	0.35	1.00	0.36

TABLE 3.29 Estimated parameters for transformation:
 $u(x) = a \ln x + b$ with maximum likelihood,
 moment and graphical methods

Month	Method					
	Maximum likelihood		Moment		Graphical	
	a	b	a	b	a	b
Jan	2.94	0.18	2.94	0.18	2.84	0.17
Feb	3.33	0.13	3.33	0.13	3.26	0.15
March	3.45	0.14	3.45	0.14	3.42	0.14
April	2.56	0.18	2.56	0.18	2.48	0.18
May	2.78	0.17	2.78	0.17	2.69	0.17
June	2.94	0.18	2.94	0.18	2.86	0.17
July	2.94	0.18	2.94	0.18	2.84	0.17
August	2.22	0.22	2.22	0.22	2.18	0.22
Sept	1.96	0.24	1.96	0.24	1.94	0.24
Oct	2.04	0.22	2.04	0.22	1.99	0.22
Nov	2.22	0.20	2.22	0.20	2.17	0.19
Dec	2.78	0.17	2.78	0.17	2.68	0.16

TABLE 3.30 Estimated parameters for transformation:
 $u(x) = a x - b/x$ with maximum likelihood,
 moment and graphical methods.

Month	Method					
	Maximum likelihood		Moment		Graphical	
	a	b	a	b	a	b
Jan	1.43	1.31	1.45	1.29	1.36	1.23
Feb	1.67	1.54	1.68	1.54	1.59	1.46
March	1.79	1.64	1.77	1.63	1.71	1.58
April	1.29	1.08	1.27	1.09	1.23	1.05
May	1.40	1.23	1.40	1.23	1.37	1.20
June	1.46	1.31	1.46	1.31	1.41	1.26
July	1.48	1.32	1.48	1.32	1.44	1.29
August	1.13	0.93	1.13	0.93	1.10	0.91
Sept	1.01	0.77	1.01	0.78	0.98	0.75
Oct	1.05	0.79	1.02	0.81	1.00	0.77
Nov	1.14	0.87	1.10	0.90	1.06	0.84
Dec	1.39	1.19	1.37	1.21	1.32	1.15

TABLE 3.31 Estimated parameters for transformation:
 $u(x) = a x^2 - b/x$ with maximum likelihood,
 moment and graphical methods.

Month	Method					
	Maximum likelihood		Moment		Graphical	
	a	b	a	b	a	b
Jan	0.64	0.85	0.69	0.71	0.57	0.66
Feb	0.86	0.90	0.90	0.91	0.77	0.83
March	1.08	1.12	1.09	1.09	1.04	1.05
April	0.73	0.78	0.75	0.74	0.71	0.72
May	0.80	0.87	0.83	0.82	0.78	0.80
June	0.74	0.87	0.78	0.79	0.68	0.74
July	0.87	0.93	0.89	0.89	0.85	0.86
August	0.59	0.69	0.61	0.61	0.56	0.60
Sept	0.53	0.61	0.56	0.55	0.52	0.54
Oct	0.59	0.61	0.59	0.57	0.55	0.56
Nov	0.69	0.67	0.69	0.66	0.66	0.64
Dec	0.84	0.86	0.86	0.84	0.81	0.81

TABLE 3.32 Estimated parameters for transformation:
 $u(x) = a\sqrt{x} - b/x$ with maximum likelihood,
 moment and graphical methods

Month	Method					
	Maximum likelihood		Moment		Graphical	
	a	b	a	b	a	b
Jan	2.06	1.80	2.06	1.80	1.98	1.73
Feb	2.31	2.09	2.30	2.09	2.20	1.99
March	2.35	2.14	2.34	2.14	2.25	2.05
April	1.69	1.39	1.66	1.40	1.58	1.32
May	1.89	1.62	1.88	1.62	1.82	1.57
June	2.00	1.75	2.00	1.76	1.93	1.69
July	2.00	1.75	2.00	1.76	1.94	1.70
August	1.59	1.24	1.54	1.23	1.50	1.19
Sept	1.37	0.99	1.33	1.01	1.27	0.95
Oct	1.39	1.01	1.33	1.02	1.26	0.95
Nov	1.43	1.08	1.38	1.10	1.27	0.99
Dec	1.80	1.52	1.77	1.53	1.67	0.43

TABLE 3.33 Estimated parameters for transformation:

$u(x) = a \ln(\ln(1+x)) + b \ln x + c \sqrt{x}$ with
maximum likelihood moment and graphical methods.

Month	Method								
	Maximum likelihood			Moment			Graphical		
	a	b	c	a	b	c	a	b	c
Jan	0.986	1.986	0.529	18.95	-14.32	7.19	21.20	-16.46	8.02
Feb	0.978	2.469	0.490	11.56	-7.16	4.42	14.43	-9.90	5.47
March	1.014	2.523	0.516	-8.36	11.08	-2.95	-10.60	13.01	-3.78
April	1.005	1.542	0.569	-10.40	11.99	-3.67	-11.51	12.95	-4.09
May	2.429	0.456	1.079	-0.135	2.80	0.13	-0.35	2.93	0.04
June	0.979	1.983	0.524	6.02	-2.61	2.40	6.48	-3.11	2.56
July	0.963	1.969	0.530	4.16	-0.94	1.71	5.18	-1.95	2.08
August	1.918	0.381	0.930	3.30	-0.88	1.44	3.49	-1.12	1.51
Sept	2.020	0.027	1.002	-3.00	4.64	-0.88	-2.42	4.05	-0.67
Oct	0.495	1.476	0.417	-7.07	8.45	-2.41	-6.81	8.15	-2.32
Nov	0.598	1.592	0.440	-17.22	18.03	-6.20	-16.75	17.55	-6.04
Dec	0.880	1.887	0.505	-16.47	17.78	-5.93	-17.53	18.68	-6.33

The fact, that large differences are observed between the parameters of the transformation eq 22, is connected of course, with the specific form of the analytical expression. The conclusion can also be drawn that the greater the number of parameters the more careful one should be in the choice of the method of estimation.

The choice of the method of estimation depends on the character of every particular problem, i.e. whether the average behaviour or the extreme events are of more interest.

3.5 Test of marginal distribution functions

The well known and frequently applied chi-square test is used as a measure of fitness of the analytical distribution functions to empirical ones. The basic concepts of this test are as follows (Cramer, 1948). The space of the variable is divided into a finite number of r intervals $S_1, \dots, S_r$ without common points, and the corresponding values of the class interval frequencies $p_1, \dots, p_r$ are set so that $p_i = p(S_i)$ and $\sum_{i=1}^r p_i = 1$. Let the corresponding class frequencies in the sample be $v_1, \dots, v_r$, so that v_i sample values belong to the interval S_i , and we have $\sum_{i=1}^r v_i = N$. A measure of the deviation of the distribution of the sample from the hypothetical distribution is:

$$\chi^2 = \sum_{i=1}^r (v_i - N p_i)^2 / N p_i \quad (3.99)$$

χ^2 is distributed in a chi-square distribution with $r-1$ degrees of freedom, for the case of a completely specified hypothetical distribution. If the parameters of the distribution are unknown, it is only necessary to reduce the number of degrees of freedom by one unit for each parameter estimated. The hypothesis to be tested is that our sample has been drawn from a population having a certain distribution $F(x, \theta_1, \dots, \theta_k)$ with some values of the k parameters $\theta_1, \dots, \theta_k$. If we find a value $\chi^2 > \chi_p^2$ where χ_p^2 is the $p\%$ value of the chi-square distribution, with $r-1-k$ degrees of freedom, the hypothesis is rejected. In the present study eight intervals with equal probability were used i.e. $r = 8$ and $p_1 = p_2 = \dots = p_8 = 0.125$.

The normal eq (3.10), two and three parameter lognormal eq (3.20) and

(3.21) and the two and three parameter gamma distributions together with the six different transformations of the unit normal distribution were tested on monthly average values of the 38 observation series presented in Table 3.1. The method of maximum likelihood was used to estimate the parameters of the first five distributions. For the six transformed normal distributions the graphical method was used to reduce the necessary computations. The chi-square test was included in the computer programs for the different distributions.

In Tables 3.34 - 3.36 the general result of the test is presented. The material was arranged in the groups of Northern rivers (region A-E) and Southern rivers (region F), and these groups were divided into the different annual periods given in part 3.3.1. The number of rejections to 10 %, 5 % and 1 % levels is given in the tables. The best results are achieved for the three parameter transformation of the unit normal distribution (eq 3.22) and the two parameter log-normal distribution. The differences from the three parameter log-normal, the two and three parameter gamma distributions and the transformation eq (3.31) are, however, not very great. Neither regional differences nor any differences between the periods can be found. It seems that the winter and beginning of snowmelt are the periods which give the most rejections for all distributions. Further, it is usually so, if the three parameter transformation or the two parameter lognormal transformation are rejected, all the other distributions are also rejected.

Table 3.34 χ^2 -test of marginal distribution functions, number of rejections on 10 % significant level

Period	Distribution	Normal	Lognormal	Gamma	Transformations of unit normal distribution										Tested total
					2 param	3 param	2 param	3 param	ax+b	a ln x+b	a x ² -b/x	a √x-b/x	ax-b/x	a ln (ln (1+x)) ² + b ln (x) + c √x	
A-E	1	61	28	42	30	43	65	33	62	34	42	33	120		
	2	18	8	8	9	10	15	6	11	9	8	4	23		
	3	8	3	7	3	3	9	5	13	14	6	0	35		
	4	25	6	9	9	9	27	5	24	6	5	7	51		
	5	25	3	10	6	10	13	4	17	18	12	4	47		
Fa,b	1	23	11	22	7	11	26	13	28	34	20	8	57		
	2	14	3	6	5	6	17	3	11	5	4	3	24		
	3	61	8	13	21	22	68	14	55	22	16	10	75		
G		9	8	13	6	10	10	11	15	14	10	5	24		
Total		244	78	130	96	124	250	94	236	156	123	74	456		

Table 3.35 χ^2 -test of marginal distribution functions, number of rejections on 5 % significant level

Period	Distribution	Normal		Lognormal		Gamma		Transformations of unit normal distribution					Tested total
		1 param	3 param	1 param	3 param	1 param	3 param	$a \ln x + b$	$a x^2 - b/x$	$a \sqrt{x} - b/x$	$a x - b/x$	$a \ln(\ln(1+x)) + b \ln(x) + c \sqrt{x}$	
A-E	1	53	15	25	21	28	57	20	57	25	21	21	120
	2	18	6	8	8	9	14	6	10	7	6	3	23
	3	5	1	6	3	2	6	2	9	11	6	0	35
	4	18	3	7	4	7	22	3	17	4	3	1	51
	5	20	3	6	5	6	16	4	14	16	7	4	47
Fa,b	1	21	10	16	3	7	21	10	18	27	14	6	57
	2	13	1	3	3	3	15	2	11	4	3	2	24
	3	57	4	6	15	15	66	5	52	10	12	7	75
G		5	8	9	4	8	4	11	11	12	8	2	24
Total		210	51	86	66	85	221	63	199	116	80	46	456

Table 3.36 χ^2 -test of marginal distribution functions, number of rejections on 1 % significant level

Distribution	Normal	Lognormal	Gamma	Transformations of unit normal distribution							Tested total		
				2 param	3 param	2 param	3 param	ax+b	a ln x+b	a x ² -b/x		a √x-b/x	ax-b/x
A-E	1	27	7	12	9	13	40	8	38	9	9	8	120
	2	14	5	5	7	7	13	5	7	3	3	1	23
	3	1	3	1	1	1	1	1	2	8	1	0	35
	4	11	0	2	1	2	12	0	9	2	0	0	51
	5	12	1	1	3	3	10	1	6	10	2	1	47
Fa,b	1	8	2	3	0	1	14	2	12	15	8	2	57
	2	8	0	0	1	2	12	0	7	1	2	1	24
	3	47	0	2	8	6	49	2	38	2	6	1	75
G	1	6	8	4	7	1	9	9	8	10	7	2	24
Total	129	24	34	34	42	152	28	127	60	38	16	456	

4. MULTIVARIATE DISTRIBUTION FUNCTIONS

Multivariate distribution functions have been applied by some authors for the description of river runoff at adjacent time periods. A fundamental problem is to describe the relation between the multivariate distribution and the corresponding marginal distributions. If the marginal variables considered, are distributed normally, we can assume that they have a joint multivariate normal distribution. As the runoff values are not normally distributed, they should be transformed first to normality. This approach is used by Kartvelishvili (1967) and Moran (1969). It can be easily generalized to the case of a vector process, i.e. runoff at several points along a river or at neighbouring rivers.

Another approach to the construction of bivariate distributions is to define their probability density by series of the type (Lancaster, 1958):

$$dF(x, y) = \{1 + \sum_{i=1}^{\infty} R_i x^{(i)} y^{(i)}\} dG(x) dH(y) \quad (4.1)$$

where $dG(x)$ and $dH(y)$ are the marginal probability density functions, $x^{(i)}$ and $y^{(i)}$ are certain sets of orthonormal functions and the R_i , $i=1, 2, \dots$ are the canonical correlations. Expansions of this kind have been given by Sarmanov O.V.(1960) and Vere-Jones (1967) for the case when the marginal distributions are gamma distributions, and for the more general case of gamma distributions with different parameters by Sarmanov I.O.(1968). The application of these bivariate gamma distributions for the description of river runoff is shown in Blochinov and Sarmanov O.V.(1968) and Sarmanov I.O.(1969). It should be noted that bivariate normal distribution can also be expressed in the form of eq (4.1).

In this part these two approaches will be discussed, as well as the procedure of the parameter estimation of the multivariate distributions.

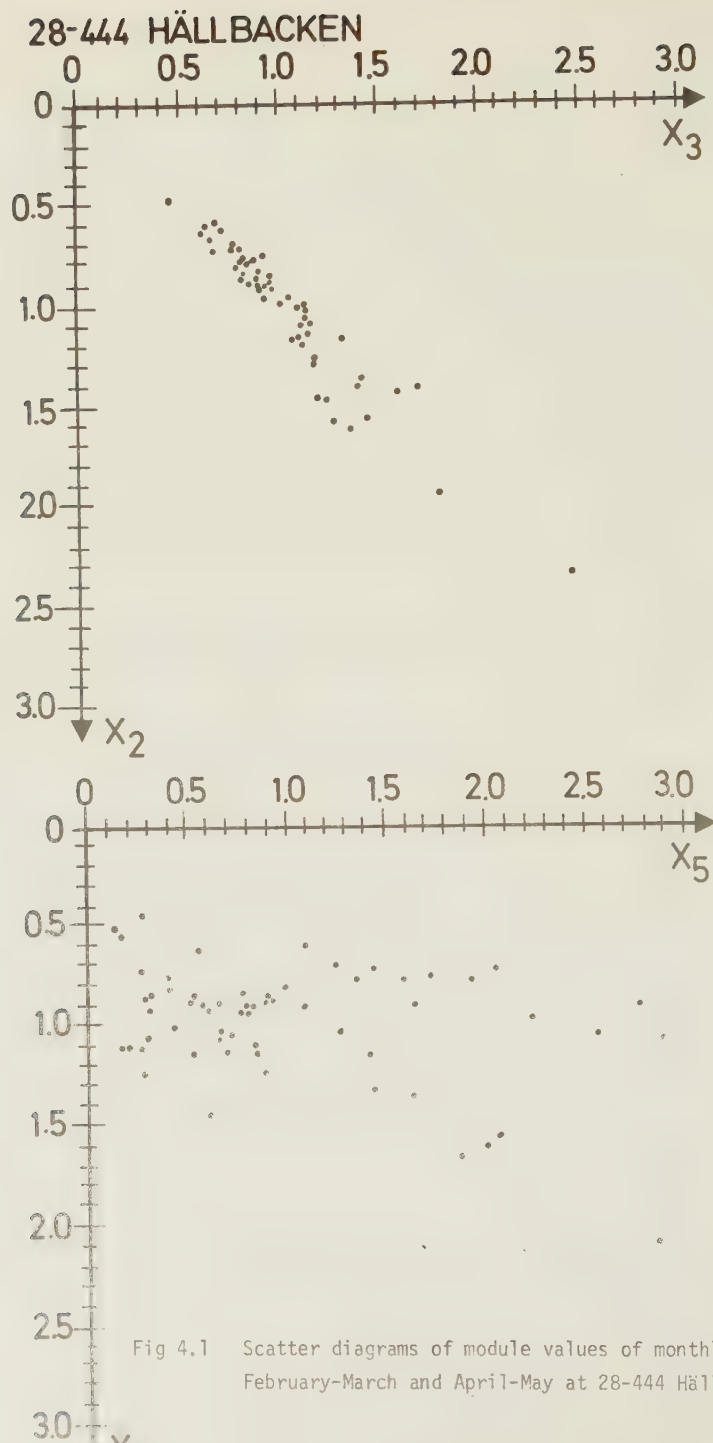


Fig 4.1 Scatter diagrams of module values of monthly runoff for February-March and April-May at 28-444 Hällbacken.

4.1 Scatter diagrams

Empirical bivariate distribution functions can be represented graphically by the so-called scatter diagrams. Such diagrams can give, of course, very rough ideas of the relation between two variables, as the number of observations is usually quite scarce. They can, however, give a visual measure of the degree of dependence between the two variables. Furthermore, they can give some knowledge about the character of this relation. The important properties of the relation in this connection are those of linearity and symmetry. The scatter diagrams can give some ideas about to which extent the given sample shows these two properties. However, these diagrams are not suitable for the precise analysis of the relations between adjacent periods of runoff.

In Figs 4.1 and 4.2 some examples of scatter diagrams are given. In the first one module values of monthly runoff for February and March at 28-444 Hällbacken are plotted. The dependence between the two months is very great and seems to be linear but not symmetric. In the other example module values from April and May are used. The dependence here is very small and almost nothing can be said about its character. Fig 4.2 shows the effect of standardization and normalization of the observations. Data are used from 4-18 Möjärv for November and December. The normalized values show symmetric and linear relation, when compared to the standardized values.

4.2 The bivariate normal distribution

The bivariate normal probability density distribution is defined by the following formula:

$$f(x_1, x_2; m_1, m_2, \sigma_1, \sigma_2, \rho) = ((2\pi)^2 (1-\rho^2) \sigma_1^2 \sigma_2^2)^{-1/2} \exp \left\{ -\left[\frac{(x_1 - m_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right] / 2(1-\rho^2) \right\} \quad (4.2)$$

4-18 Mojärv

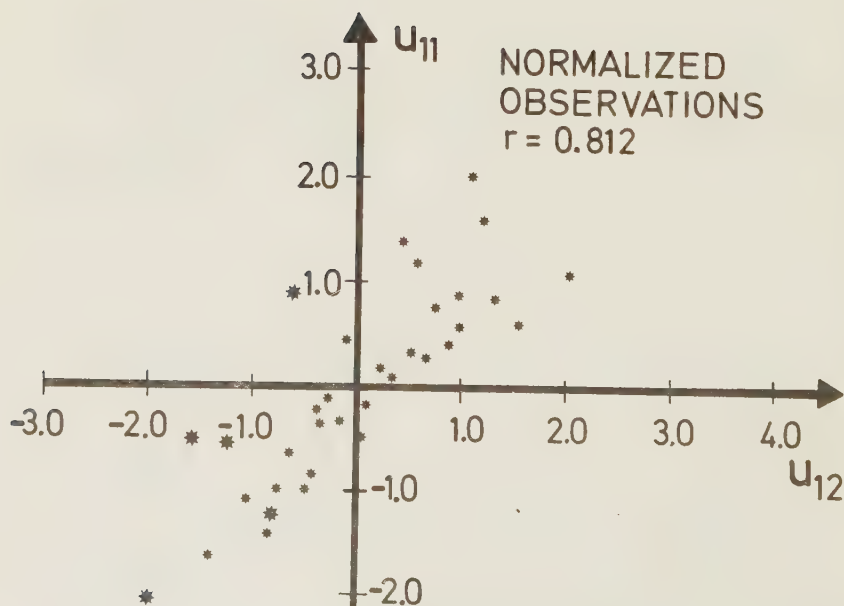
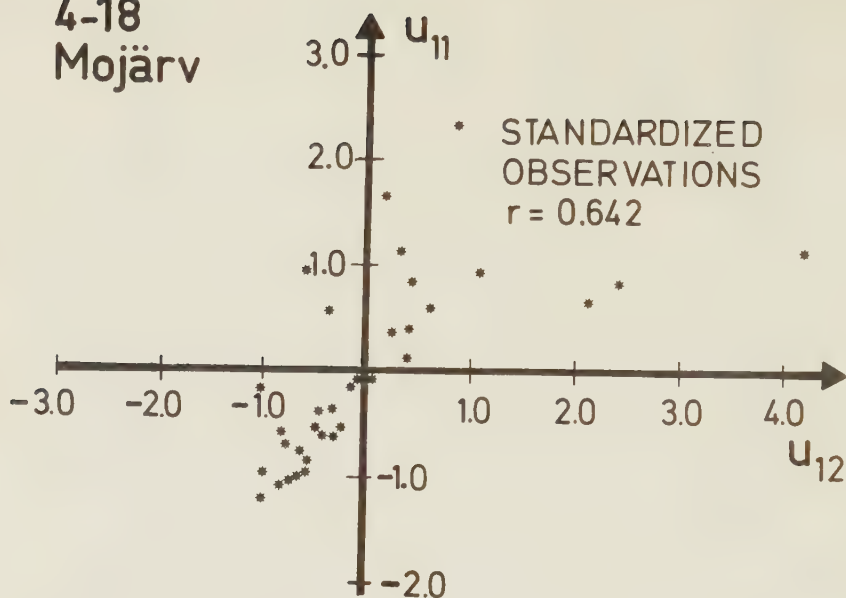


Fig 4.2 Scatter diagrams of standardized and normalized module values of monthly river runoff for November-December at 4-18 Mojärv.

The dependence between the two variables x_1 and x_2 is thus fully described by one parameter ρ , the coefficient of correlation. The dependence is linear and symmetric, which is illustrated in Fig 4.3. The parameters to the two bivariate normal distributions have been calculated from the data represented in Fig 4.1.

Let us now use some functions $u_1(x_1)$ and $u_2(x_2)$ instead of x_1 and x_2 . We assume that these new variables are bivariate normally distributed. We further use the standardized form of eq (4.2). The density functions for the new variables u_1 and u_2 will have the expression:

$$f(u_1, u_2; \rho) = ((2\pi)^2 (1-\rho^2))^{-1/2} \exp\{-(u_1^2 - 2\rho u_1 u_2 + u_2^2) / 2(1-\rho^2)\} \quad (4.3)$$

where ρ now is the correlation coefficient between the new variables. The density function for the variables x_1 and x_2 will now have the form:

$$f(x_1, x_2; \rho) = ((2\pi)^2 (1-\rho^2))^{-1/2} \exp\{-(u_1^2(x_1) - 2\rho u_1(x_1) u_2(x_2) + u_2^2(x_2)) / 2(1-\rho^2)\} du_1 / dx_1 du_2 / dx_2 \quad (4.4)$$

In dependence on the form of the transformation functions u_1 and u_2 , formula 4.4 can describe nonlinear and nonsymmetric relations between the variables x_1 and x_2 . Johnson (1949b) has suggested a system of bivariate distributions based on the bivariate normal distribution and the transformation discussed above in part 3.4.1. In Fig 4.4 two examples are shown, where the logarithmic transformation eq (3.30) of bivariate distributions have been used. The data for parameter estimation is the same as the one represented in Fig 4.1.

If the variables x_1 and x_2 are distributed according to the bivariate normal distribution given by eq 4.2, it can be shown that the marginal distributions of the variables x_1 and x_2 are normal with the parameters m_1 , σ_1 and m_2 , σ_2 , respectively (see, for instance,

28-444 Hällbacken

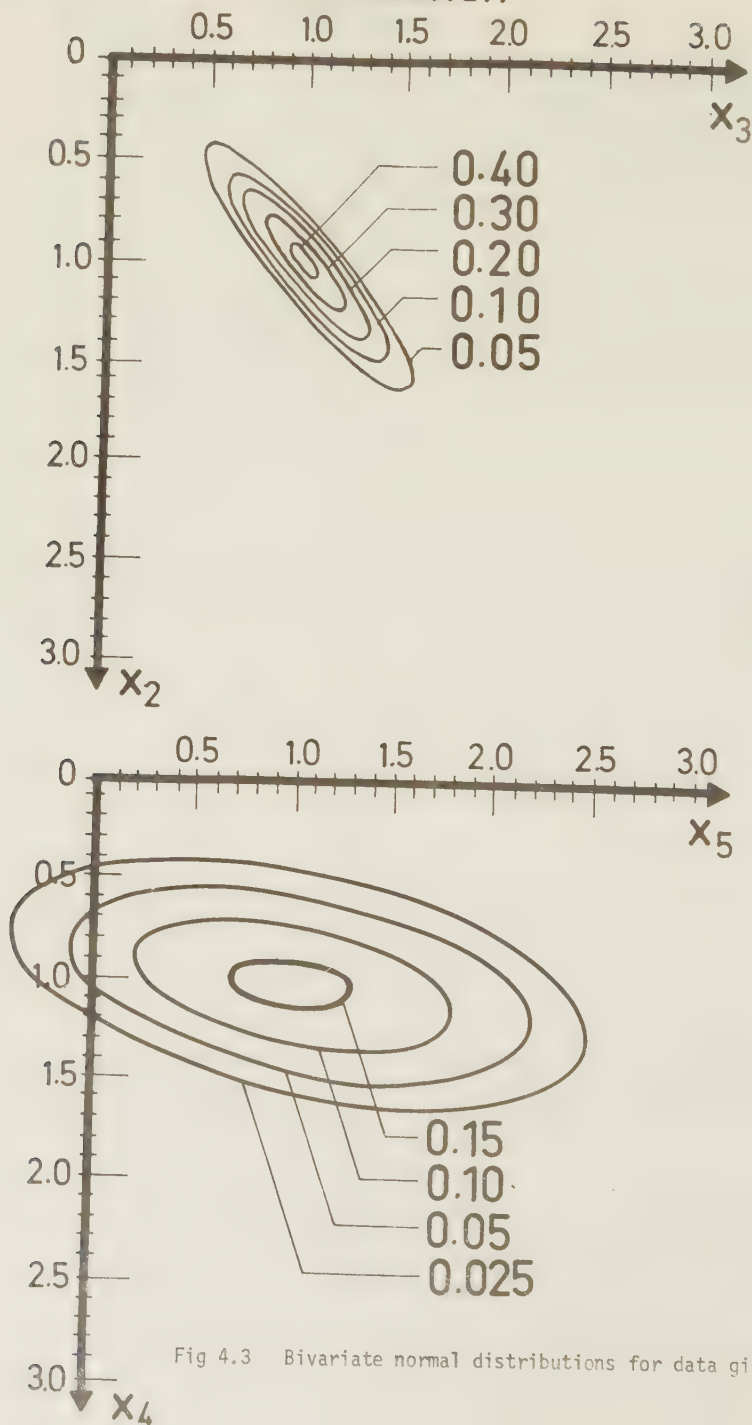


Fig 4.3 Bivariate normal distributions for data given in Fig 4.1.

Ventsel, 1964). Assuming that transformations u_1 and u_2 have a joint distribution given by eq 4.3 we similarly have that u_1 and u_2 are marginally unit normal distributed. We shall further assume that the reverse relations are valid i.e. if the marginal variables x_1 and x_2 are normally distributed with the parameters m_1 , σ_1 and m_2 , σ_2 , respectively, then the joint bivariate distribution is given by eq (4.2). The similar assumption is made for u_1 and u_2 .

4.3 The bivariate gamma distribution

A bivariate gamma distribution is defined by the following density function (Sarmanov (1960), Vere Jones (1967)):

$$f(x_1, x_2; \gamma, R) = f(x_1; \gamma) f(x_2; \gamma) \sum_{k=0}^{\infty} L_k^{\gamma-1}(x_1) L_k^{\gamma-1}(x_2) \cdot R^k$$

$$0 \leq x_1, x_2 < \infty, \gamma > 0, 0 \leq R < 1 \quad (4.5)$$

here $f(x_i, \gamma)$, $i = 1, 2$ are the one parameter gamma distributions eq (3.14) and $L_k^{\gamma-1}(x_i)$, $i = 1, 2$ are the generalized orthonormal Laquerre polynomials with the weight function $f(x_i, \gamma)$.

The bivariate gamma distribution above can be easily generalized to the case of marginal two parameter gamma distributions. Using the two parameter gamma distribution eq (3.34) we derive the following formula:

$$f(x_1, x_2; \gamma, \beta, R) = f(x_1; \gamma, \beta) f(x_2; \gamma, \beta)$$

$$\sum_{k=0}^{\infty} L_k^{\gamma-1}(x_2/\beta) L_k^{\gamma-1}(x_1/\beta) R^k \quad (4.6)$$

The conditional mean value for the distribution function above, is given by:

$$m_{x_2} | x_1 = 1 + R (x_1 - 1)$$

and the conditional standard deviation:

$$\sigma_{x_2}^2 | x_1 = \sigma \sqrt{(1-R)^2 + 2R(1-R)x_1}$$

28-444 Hällbacken

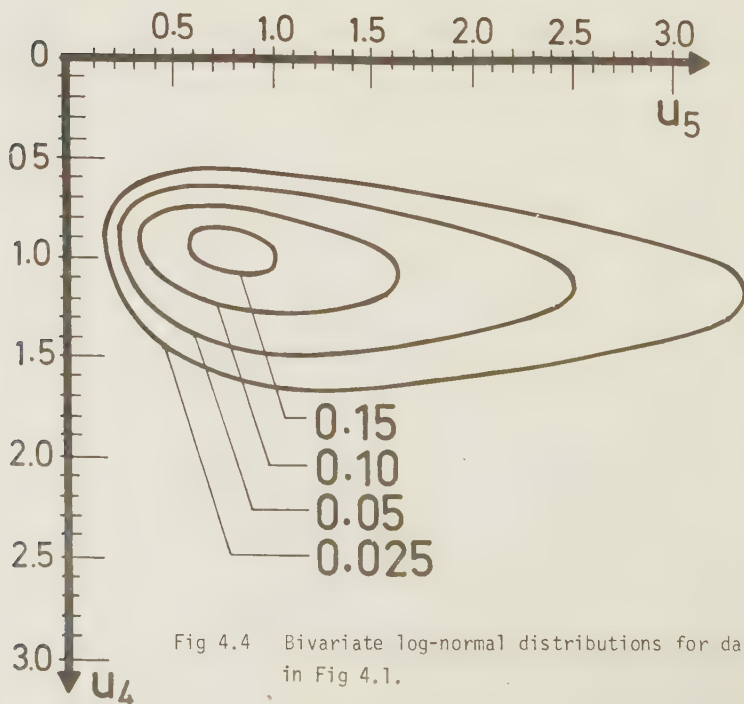
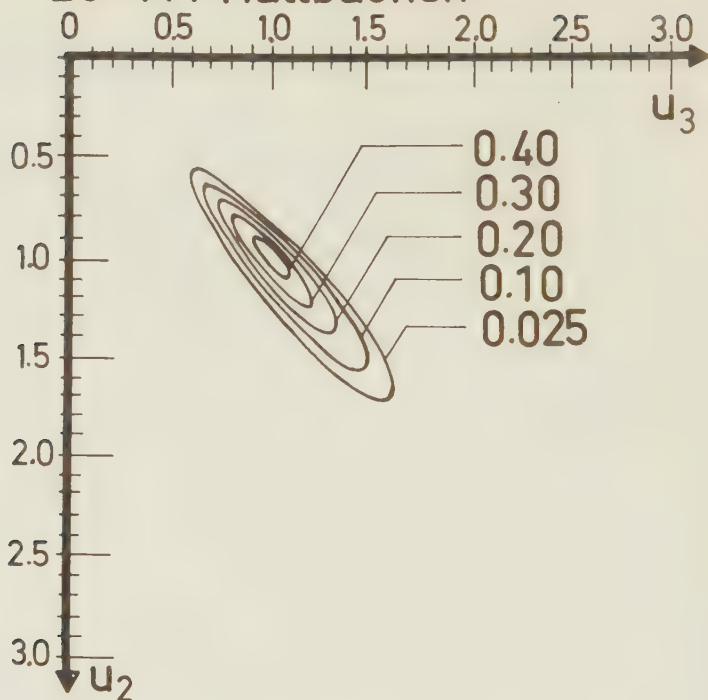


Fig 4.4 Bivariate log-normal distributions for data given in Fig 4.1.

Equation (4.6) has the following three properties (Blochinov and Sarmanov, 1968):

1. The dependence between x_1 and x_2 is linear (cf. the conditional mean).
2. If the canonical correlation coefficient R is zero, the variables x_1 and x_2 are independent.
3. The canonical correlation coefficient is nonnegative ($R \geq 0$).

A generalization of the distribution (4.6) has been suggested by Sarmanov I O (1968), for the case, when the marginal distribution functions have different parameters. The generalized bivariate gamma density distribution is defined by:

$$f(x_1, x_2; \gamma_1, \beta_1, \gamma_2, \beta_2, \{c_k\}) = f(x_1; \gamma_1, \beta_1) f(x_2; \gamma_2, \beta_2) \sum_{k=0}^{\infty} c_k L_k^{\gamma_1-1}(x_1 / \beta_1) L_k^{\gamma_2-1}(x_2 / \beta_2) \quad (4.7)$$

here $\{c_k\}$ is a series of non-negative numbers, which satisfies the condition $\sum_{k=1}^{\infty} c_k^2 < \infty$. The other notations were already explained above. The coefficients c_k , $k = 1, 2, \dots$ are the correlation coefficients between $L_k^{\gamma_1-1}(x_1)$ and $L_k^{\gamma_2-1}(x_2)$, and are determined by the following expression (Sarmanov I O, 1969):

$$c_k = \lambda^k \Gamma(\delta+k) / \Gamma(\delta) \cdot \sqrt{\Gamma(\gamma_1)\Gamma(\gamma_2)} / \sqrt{\Gamma(\gamma_1+k)\Gamma(\gamma_2+k)} \quad (4.8)$$

where λ and δ are parameters. The total number of parameters of the distribution (4.7) should thus be six.

Moran (1969) suggests a bivariate gamma distribution based on the bivariate unit normal distribution, where the variables are transformed to a two parameter gamma distribution. The joint probability density distribution is thus given by the formula:

$$\begin{aligned}
 f(x_1, x_2; \gamma_1, \beta_1, \gamma_2, \beta_2, \rho) = \\
 (2\pi)^{-1} (1-\rho^2)^{-1/2} \exp \{ - (2(1-\rho^2))^{-1} \\
 \{ \phi^{-1} \{ F(x_1; \beta_1, \gamma_1) \} \}^2 - 2\rho \phi^{-1} \{ F(x_1; \beta_1, \gamma_1) \} \\
 \phi^{-1} \{ F(x_2; \beta_2, \gamma_2) \} + \{ \phi^{-1} \{ F(x_2; \beta_2, \gamma_2) \} \}^2 \} \\
 \cdot x_1^{\gamma_1-1} x_2^{\gamma_2-1} \beta_1^{-\gamma_1} \beta_2^{-\gamma_2} \quad (4.9)
 \end{aligned}$$

4.4 The multivariate normal distribution

Let u_i , $i = 1, 2, \dots, 12$ represent some transformation of the module of runoff for the i -th month. The u_i has a marginal unit normal distribution. We shall assume that u_i , $i = k, k-1, \dots, k-n$ have a joint multivariate distribution:

$$\begin{aligned}
 f(u_k, u_{k-1}, \dots, u_{k-n}) = \\
 = ((2\pi)^{n+1} D)^{-1/2} \exp \{ (2D)^{-1} \\
 \begin{vmatrix}
 1 & \rho_{k, k-1} & \rho_{k, k-n} & u_k \\
 \rho_{k-1, k} & 1 & \dots & \rho_{k-1, k-n} u_{k-1} \\
 \vdots & \vdots & \ddots & \vdots \\
 \rho_{k-n, k} & \rho_{k-n, k-1} & \dots & 1 & u_{k-n} \\
 u_k & u_{k-1} & u_{k-n} & 0
 \end{vmatrix}
 \end{vmatrix} \quad (4.10)
 \end{aligned}$$

where

$$D = \begin{vmatrix}
 1 & \rho_{k, k-1} & \dots & \rho_{k, k-n} \\
 \rho_{k-1, k} & 1 & \dots & \rho_{k-1, k-n} \\
 \vdots & \vdots & \ddots & \vdots \\
 \rho_{k-n, k} & \rho_{k-n, k-1} & \dots & 1
 \end{vmatrix}$$

and ρ_{ij} , $i, j = k, k-1; \dots, k-n$ are the correlation coefficients between u_i and u_j .

Let us now consider a m -dimensional vector $\vec{u}_i = (u_i^1, u_i^2, \dots, u_i^m)$ i.e. transformed monthly module values at several points in one river or in the neighbouring rivers for the month i . The marginal distribution of this vector is:

$$f(\vec{u}_i) = ((2\pi)^m \det \Delta_{i \ i})^{-1/2} \exp \{ (2 \det (\Delta_{i \ i}))^{-1} \begin{vmatrix} \Delta_{i \ i} & u_i^T \\ u_i & 0 \end{vmatrix} \} \quad (4.11)$$

where

$$\Delta_{i \ i} = \begin{pmatrix} 1 & \rho_{i \ i}^{1 \ 2} & \dots & \rho_{i \ i}^{1 \ m} \\ \rho_{i \ i}^{2 \ 1} & 1 & & \rho_{i \ i}^{2 \ m} \\ \vdots & & \ddots & \vdots \\ \rho_{i \ i}^{m \ 1} & \rho_{i \ i}^{m \ 2} & \dots & 1 \end{pmatrix}$$

$$\vec{u}_i = (u_i^1, u_i^2, \dots, u_i^m)$$

$$\vec{u}_i^T = \begin{pmatrix} u_i^1 \\ u_i^2 \\ \vdots \\ u_i^m \end{pmatrix}$$

and $\rho_{i \ i}^{s \ t}$; $s, t=1, \dots, m$ are the correlation coefficients between the variables u_i^s and u_i^t at the i -th month. The multivariate distribution for the vectors $\vec{u}_i$, $i = k, k+1, \dots, k+n$ is now defined by the following probability density distribution function:

$$\begin{aligned} f(u_k^1, u_k^2, \dots, u_k^m, u_{k+1}^1, \dots, u_{k+1}^m, \dots, u_{k+n}^1, \dots, u_{k+n}^m) = \\ = (2\pi)^{\frac{(n+1)m}{2}} D^{-1/2} \exp \{ -1 / (2 D) D_u \} \end{aligned} \quad (4.12)$$

where

$$D_u = \begin{vmatrix} \Delta_{kk} & \Delta_{k,k-1} & \dots & \Delta_{k,k-n} & \vec{u}_k^T \\ \Delta_{k-1,k} & \Delta_{k-1,k-1} & \dots & \Delta_{k-1,k-n} & \vec{u}_{k-1}^T \\ \vdots & \vdots & & \vdots & \vdots \\ \Delta_{k-n,k} & \Delta_{k-n,k-1} & \dots & \Delta_{k-n,k-n} & \vec{u}_{k-n}^T \\ \vec{u}_k & \vec{u}_{k-1} & \dots & \vec{u}_{k-n} & 0 \end{vmatrix}$$

$$D = \begin{vmatrix} \Delta_{kk} & \Delta_{k,k-1} & \dots & \Delta_{k,k-n} \\ \Delta_{k-1,k} & \Delta_{k-1,k-1} & \dots & \Delta_{k-1,k-n} \\ \vdots & \vdots & & \vdots \\ \Delta_{k-n,k} & \Delta_{k-n,k-1} & \dots & \Delta_{k-n,k-n} \end{vmatrix}$$

$$\Delta_{i,j} = \begin{pmatrix} \rho_{ij}^{11} & \rho_{ij}^{12} & \dots & \rho_{ij}^{1m} \\ \rho_{ij}^{21} & \rho_{ij}^{22} & \dots & \rho_{ij}^{2m} \\ \vdots & \vdots & & \vdots \\ \rho_{ij}^{m1} & \rho_{ij}^{m2} & \dots & \rho_{ij}^{mm} \end{pmatrix} \quad i, j = k, k-1, \dots, k-n$$

and ρ_{ij}^{st} ; $i, j = k, k-1, \dots, k-n$; $s, t = 1, \dots, m$ are the correlation coefficients between u_s^i and u_t^j .

4.5 Parameter estimation

To determine the parameters of the bivariate and multivariate distributions referred to above, the same methods as those discussed in part 3.4.5 are, in principle, available. The theoretical and numeri-

cal calculations will be, however, be much complicated, when performing a joint derivation of all parameters of respective distributions. Not only the specific parameters that describe the dependence between the variables, i.e. the coefficient of correlation and the canonical correlation coefficients, should be calculated for the normal and gamma distributions, respectively, but also the parameters of the marginal distributions. If we use the maximum likelihood method to estimate the parameters of the bivariate normal distribution eq (4.2) it comes out that the marginal empirical distributions contain all the information about the parameters of the marginal distributions. When using other than linear transformations of unit normal distribution this is not the case. The formulas for the estimation of the parameters of the marginal distribution for some certain month involve thus observations from adjacent months. It is difficult to foretell theoretically the amount of extra information, that can be obtained from these. In a discussion of parameter estimates of the distribution equation (4.9) Moran (1969) assumes that the extra information is vanishingly small so that estimates calculated with the maximum likelihood method for the marginal distribution should be asymptotically efficient. In any case, it is possible to guess that the marginal estimates will be close to the estimates from the joint distribution, as one can assume that the main part of the information about the marginal parameters should be contained in the marginal sample.

In this study the marginal sample alone has been used to calculate the parameters of the marginal distributions. The way to estimate the coefficient of correlation of the normal distribution and the canonical correlations of the bivariate gamma distribution will be discussed below.

4.5.1 The coefficient of correlation

The coefficient of correlation ρ_{12} between two random variables X_1 and X_2 is defined by:

$$\rho_{12} = E \{ (x_1 - E(x_1)) (x_2 - E(x_2)) \} / \{ E(x_1 - E(x_1))^2 E(x_2 - E(x_2))^2 \}^{-1/2} \quad (4.12)$$

The correlation coefficient is a measure of the degree of linearity shown by the joint probability distribution of X_1 and X_2 . It can be shown that the parameter ρ in the expression for the bivariate normal distribution eq (4.3) is equal to the coefficient of correlation (see, for instance, Ventse1 1964). In the same way the dependence between the unit normal variables of an multivariate normal distribution eq (4.10) is described by the $(n+1) n / 2$ coefficients of correlation ρ_{ij} ; $i=k, k-1, \dots, k-n$; $j = k-1, k-2, \dots, k-n$. For the distributions other than normal one cannot expect that the coefficient of correlation gives a full description of the dependence between two random variables.

The coefficient of correlation of the population ρ_{ij} can be approximated by the coefficient of correlation of the sample r_{ij} given by:

$$r_{ij} = 1 / N \sum_{k=1}^N (x_i^k - \bar{x}_i) (x_j^k - \bar{x}_j) / (1 / N \sum_{k=1}^N (x_i^k - \bar{x}_i)^2 \sum_{k=1}^N (x_j^k - \bar{x}_j)^2)^{1/2} \quad (4.13)$$

The coefficients of correlation of the transformed multivariate distributions could be also calculated from a joint derivation by the method of maximum likelihood. These do not necessary coincide with the sample correlation coefficients. Such a derivation, however, would be very complicated. A simplification is suggested by Kartvelishvili (1967). It is based on the distributions of pairs of components x_i and x_j ; $i=1, \dots, 12$; $j=i, \dots, 12$. We have to assume here that the dependence does not extend beyond a period of six months.

We shall calculate the parameters of the marginal distribution functions of the multivariate distribution functions (eqs 4.10, 12) from the marginal samples. We shall thus assume that the transformed variables have a multivariate unit normal distribution. To estimate the correlation coefficients between these transformed variables, we shall use the expression for the sample correlation coefficient

given by eq 4.13, but change in the formula x_i and x_j to $u_1(x_i)$ and $u_2(x_j)$. The correlation coefficients calculated in this way are, of course, dependent on the transformations u_1 and u_2 , and differ in general from the correlation coefficients of the original sample. A good transformation function should naturally normalize the data but also linearize the relation between the two variables. This will give rise to an increase of the coefficient of correlation, which implies a gain of information.

Two examples of correlation coefficients calculated from standardized and normalized observations, respectively, are shown in Figs 4.5 and 4.6. To normalize the data the empirical transformation function (cf 3.4.2) has been used directly. It can be seen that the correlation coefficient is in general higher for the normalized values. The difference is in most cases small but can be as high as 0.2. It is difficult to find any regular pattern when these large differences are observed. They are, however, usually connected with high skewness of at least one of the marginal empirical distributions. The effect of normalizing of observations was shown in scatter diagrams in Fig 4.2. For the example shown there, the increase in the correlation coefficient was from 0.642 to 0.812.

4.5.2 Regional variations in the coefficient of correlation

In this part we shall use the notation $r_{i,k}$, where $k = (j-i)$, instead of r_{ij} for the correlation coefficient between month i and j .

The coefficient of correlation for monthly runoff values, shows an annual periodic variation in the same way as the empirical marginal moments (cf 3.3.1). The periodic variations are most apparent for the northern groups of rivers (A-E, cf 3.3). The general patterns of $r_{i,1}$; $i=1, \dots, 12$ are more or less the same for these groups (Fig 4.5); a very high coefficient of correlation during winter, very low or negative during the snowmelt period and afterwards a gradual increase again to the high winter values. The same variations can be found for $r_{i,2}$ and $r_{i,3}$, though not that well pronounced. Climatic runoff forming factors seem to explain most of the variations in the correlation coefficient of monthly runoff. Snowmelt

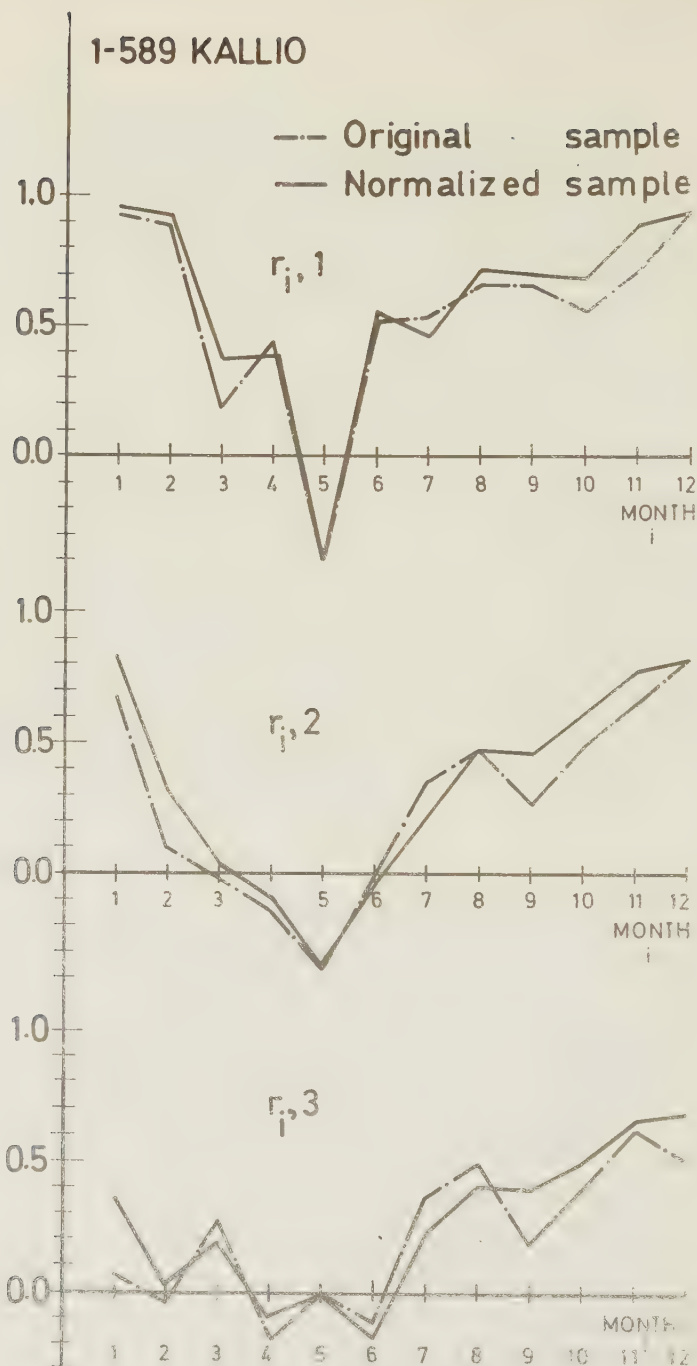


Fig 4.5 Annual variation of first to third order correlation coefficients of standardized and of normalized data for monthly river runoff at 1-589 Kallio.

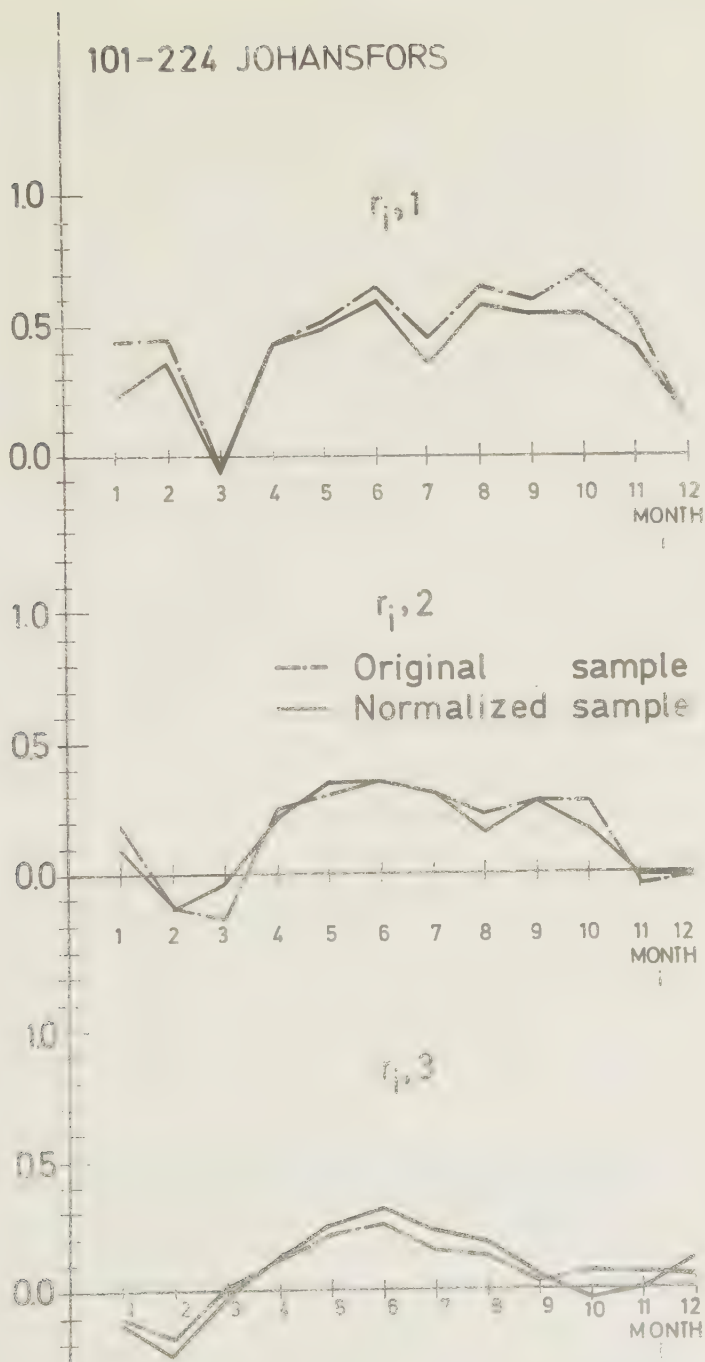


Fig 4.6 Annual variation of first to third order correlation coefficients of standardized and of normalized data for monthly river runoff at 101-224 Johansfors

in spring and rainfall in summer and autumn are processes with almost no or relatively low coefficient of correlation. Correlation caused by carry-over from one month to another due to storage in the ground-water zone, lakes and rivers, is apparently of less importance while dealing with monthly runoff, but is, of course, also included in the total correlation coefficient. One should, for instance, expect that for very large catchments the amplitude in the annual periodic variation of the correlation should be smaller than for small catchments. No such regularities could be found, except for the outflow from the two very large lakes which were included in the investigation (Fig 4.7). This may depend on the fact that the observation material is too small and that other physiographic factors have influence. Using smaller time periods than one month, we can expect less variations in the coefficient of correlation as the relative influence of the storage carry-over should be larger.

For the southern groups of rivers the variations are smaller and no significant annual periodicity seems to be present. (Fig 4.7). A decrease in the correlation coefficient at the beginning of snow-melt (March) can, however, be traced. The high values that were observed during winter for the northern rivers were not found. The climatic runoff forming factors are less stable during winter in this region. Periods of snowmelt can occur at any time.

From the above discussion we can draw the conclusion, that in most cases the correlograms of monthly runoff are functions of time. The monthly runoff process therefore cannot be considered to be stationary, even when the instationarity in the first, second and maybe also third order marginal moments have been removed from the data, for instance, by some two or three parameter transformation, with the parameters calculated from these moments.

Two examples of correlograms for different months are shown in Figs 4.8-9, one from the northern region (28-444 Hällbacken) and one from the southern region (74-177 Järnforsen). The correlograms are also drawn for negative time lags. For a stationary series the correlograms should be symmetric around the origin, which is hardly the case.

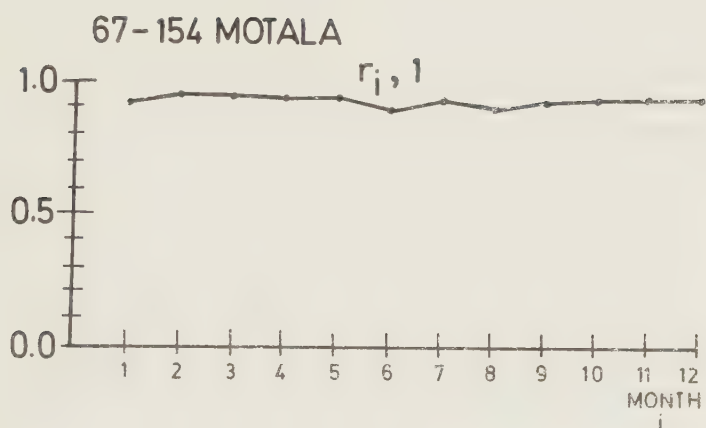


Fig 4.7 Annual variation of first order correlation coefficient for monthly river runoff at 67-154 Motala.

The correlation coefficients between different observation points along a river or between observation points on neighbouring rivers are usually large. The distance between these points should not be too large, of course, and they should be situated in the same physiographic area. In table 4.1 correlation matrices for the three observation points on the river Torne älv, number 1 on the map (Fig 3.4), are shown. The low correlation coefficients between the observation point 1-589, situated in the low land and the other two points are noticeable. This supports the suggestion to divide the northern hydrological regions into one mountain and one lowland area. Another example from the south of Sweden is shown in table 4.2. The correlation here is also very high. The space correlation between different observation points should be studied further, in connection with a more thorough analysis of hydrological regions of Sweden.

The information, that one can get from a correlogram should be, however, handled with care for several reasons. The value of $r_{i,k}$ is subjected to sampling errors that increase with k and decrease with the number of observations. For a stationary process the variance of $r_{i,k}$ which is actually not dependent on i , can be determined (see, for instance, Kendall, 1973). It is noticeable, that besides the dependence on k , the variance is larger for larger values of the population correlation coefficient. The number of observations of river runoff is relatively small and, therefore, the errors in $r_{i,k}$ can be significant, especially for large k . This can lead to the fact that the empirical correlograms do not damp out with k , as one could expect from the theory (cf Fig 4.8).

One should take into consideration, while analysing correlograms, that successive values of $r_{i,k}$ for both i and k are correlated, and maybe highly so (Bartlett, 1935). Yevjevich (1972a) draws attention to spurious correlation between successive correlation coefficients caused by computation procedure.

$r_{i,k}$ is a biased estimator of $\rho_{i,k}$. For normal stationary series the biases are downwards and of order $1/N$ (Marriott and Pope 1954, Kendall 1954). Several formulas have been suggested to eliminate biases of this order (Kritskij and Menkel, 1946, Kendall, 1954, Quenulle, 1956). These correctors can, however, lead to less

28-444 HÄLLBACKEN

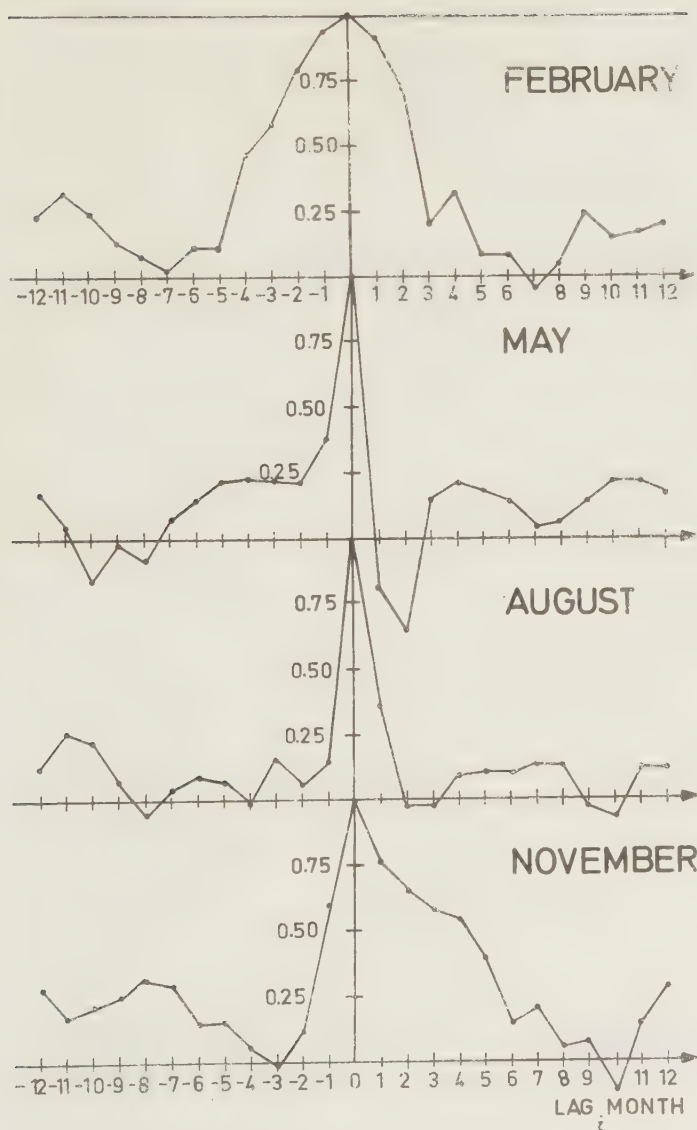


Fig 4.8 Monthly correlograms for some different months for 28-444 Hällbacken.

74-177 JÄRNFORSSEN

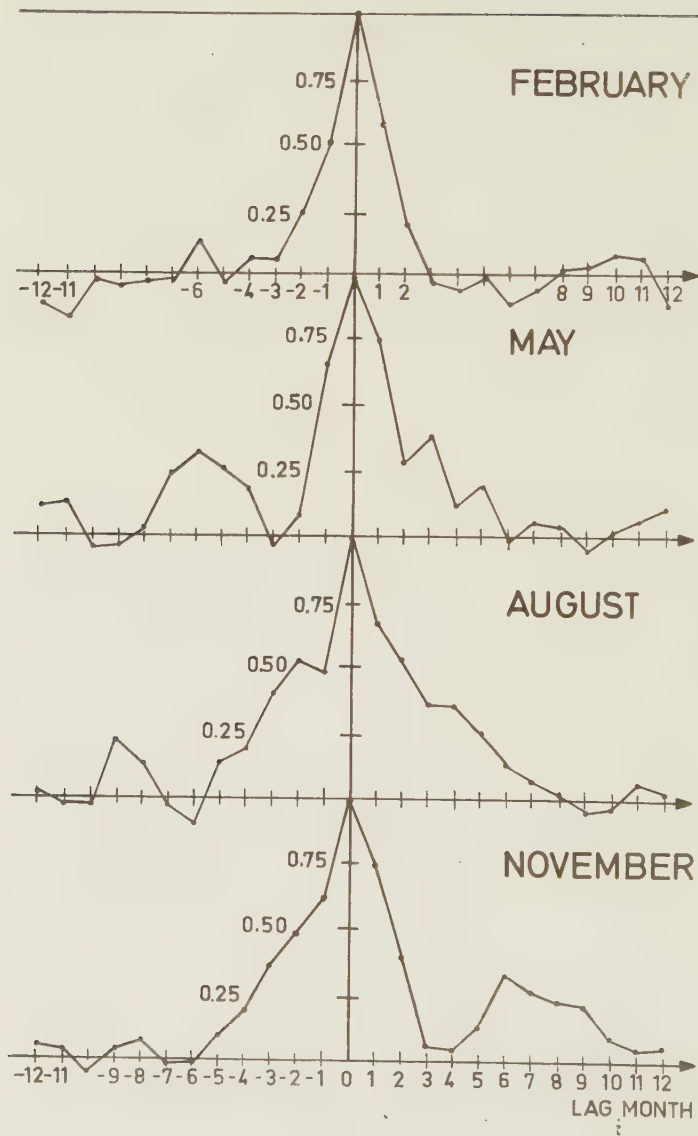


Fig 4.9 Monthly correlograms for some different months for 74-177 Järnforsen.

TABLE 4.1 Space correlation matrices for different months
for the observation points 1-959 Abiskojokk,
1-3 Jukkasjärvi and 1-589 Kallio.

January:			February:			March:		
1.0	0.674	0.039	1.0	0.705	0.032	1.0	0.575	-0.146
0.674	1.0	0.328	0.705	1.0	0.295	0.575	1.0	0.250
0.039	0.328	1.0	-0.032	0.295	1.0	-0.146	0.250	1.0
April:			May:			June:		
1.0	0.546	0.132	1.0	0.831	0.732	1.0	0.810	-0.137
0.546	1.0	0.402	0.831	1.0	0.843	0.810	1.0	0.309
0.132	0.402	1.0	0.732	0.843	1.000	-0.137	0.309	1.0
July:			August:			September:		
1.0	0.896	0.497	1.0	0.877	0.558	1.0	0.819	0.626
0.896	1.0	0.638	0.877	1.0	0.721	0.819	1.0	0.724
0.497	0.638	1.0	0.558	0.721	1.0	0.626	0.724	1.0
October:			November:			December:		
1.0	0.756	0.278	1.0	0.828	0.303	1.0	0.838	0.124
0.756	1.0	0.512	0.828	1.0	0.548	0.838	1.0	0.400
0.278	0.512	1.0	0.303	0.548	1.0	0.124	0.400	1.0

TABLE 4.2 Space correlation matrices for different months
for the observation points 101-224 Johansfors,
103-974 Kila and 105-227 Asbro.

January:			February:			March:		
1.0	0.915	0.868	1.0	0.956	0.934	1.0	0.963	0.906
0.915	1.0	0.869	0.956	1.0	0.937	0.963	1.0	0.925
0.868	0.869	1.0	0.934	0.937	1.0	0.906	0.925	1.0
April:			May:			June:		
1.0	0.928	0.886	1.0	0.932	0.677	1.0	0.930	0.788
0.928	1.0	0.831	0.932	1.0	0.745	0.930	1.0	0.854
0.886	0.831	1.0	0.677	0.745	1.0	0.788	0.854	1.0
July:			August:			September:		
1.0	0.895	0.891	1.0	0.841	0.887	1.0	0.941	0.947
0.895	1.0	0.867	0.841	1.0	0.882	0.941	1.0	0.947
0.891	0.867	1.0	0.857	0.882	1.0	0.922	0.947	1.0
October:			November:			December:		
1.0	0.932	0.954	1.0	0.966	0.952	1.0	0.975	0.852
0.932	1.0	0.945	0.966	1.0	0.982	0.975	1.0	0.875
0.954	0.945	1.0	0.952	0.982	1.0	0.852	0.875	1.0

accuracy in the estimates, which has been shown by Wallis and O'Connell (1972) with the help of Monte-Carlo methods for the estimation of biases and standard errors in the first order correlation coefficient.

As far as the effect of non-normality is concerned, one can expect that the biases will be at least as large as for the case of normality, probably larger. Reznikovskij (1969) also with the help of Monte-Carlo methods, but generating correlated two parameter gamma distributed samples, shows the bias and the standard deviation of correlation coefficients, calculated from such synthetic samples. The biases are in general larger than those given by Wallis and O'Connell, and are dependent on the coefficient of variation. This also implies dependence on the skew coefficient as $C_S = 2 C_V$ for the chosen distribution. Somewhat amazing result is that the bias decreases with increasing variation coefficient. An empirical formula is given for this dependence. The author also discusses the standard error of the first order serial correlation coefficient and compares it to the standard error of the correlation coefficient for a bivariate normal sample. Blochinov and Sarmanov (1968) have derived the following expression for the case of a bivariate gamma distributed sample from the distribution (4.5):

$$\sigma_Y(R) = (1-R^2) ((1+R(4+R) C_V^2 / (1+R)^2) / N)^{1/2} \quad (4.14)$$

The standard error is thus larger than in the case of a normal sample and increases with C_V .

The information we want to gain from the correlogram, besides the individual ordinates of it, is the type of model to use i.e. the order of the Markov process (cf. eq 2.13). This is more connected with the correlogram as a whole than with individual values. For a certain order of the process, all correlation coefficients with lags larger than the order, are determined by the correlation coefficients with lags less or equal to the order (cf eq 2.6). The way to determine the order will be shown in part 5 after the description of the finite difference representation of multivariate distributions.

4.5.3 Parameters of the bivariate Γ -distribution

The three parameters γ , β and R of the symmetric bivariate gamma distribution eq (4.6) can be easily determined by the method of moments. The canonical correlation coefficient R is equal to the ordinary correlation coefficient and can be calculated from formula eq (4.13) and β and γ are calculated from the relations $C_V = 1 / \sqrt{\gamma}$ and $\bar{x} = \beta \gamma$ (cf 3.4.4).

The nonsymmetric bivariate gamma distribution eq (4.7) has six parameters γ_1 , β_1 , γ_2 , β_2 , λ and δ . The parameters of the marginal distributions can be calculated from the marginal samples by the method of moments. To determine λ and δ we have to calculate the two coefficients c_1 and c_2 , which are the correlation coefficients between the generalized Lagurre functions

$$L_1^{\gamma_i-1}(x_i / \beta_i) ; i = 1, 2$$

and

$$L_2^{\gamma_i-1}(x_i / \beta_i) ; i = 1, 2$$

respectively. These functions have the formulas:

$$\begin{aligned} L_1^{\gamma_i-1}(x_i / \beta_i) &= \sqrt{\gamma_i} (1 - x_i / (\gamma_i \beta_i)) ; i = 1, 2 \\ L_2^{\gamma_i-1}(x_i / \beta_i) &= (\gamma_i (\gamma_i + 1) / 2!)^{1/2} (1 - 2 x_i / (\gamma_i \beta_i) \\ &+ x_i^2 / (\beta_i^2 \gamma_i (\gamma_i + 1))) ; i = 1, 2 \end{aligned}$$

Expressing the parameters γ_i and β_i , $i = 1, 2$ by the marginal moments (cf 3.4.4) we can get the searched coefficients as given by (Sarmanov I O 1969):

$$\hat{\gamma}_1 = (\bar{x}_1)^2 / m_{20}$$

$$\hat{\beta}_1 = m_{20} / \bar{x}_1$$

$$\hat{\gamma}_2 = (\bar{x}_2)^2 / m_{02}$$

$$\hat{\beta}_1 = m_{02} / \bar{x}_2$$

$$c_1 = m_{11} / \sqrt{m_{20} m_{02}}$$

$$c_2 = \sqrt{m_{20} m_{02}} \bar{x}_1 \bar{x}_2 / (2m_{20} m_{02}) (4 m_{11} / (\bar{x}_1 \bar{x}_2) - 2 m_{12} / (\bar{x}_1 m_{02}) - 2 m_{21} / (\bar{x}_2 m_{20}) + m_{22} / (m_{20} m_{02}) - 1)$$

$$\delta = (\sqrt{m_{20} m_{02}} / (\bar{x}_1 \bar{x}_2) c_2 / c_1^2 - 1)^{-1}$$

$$\alpha = (\sqrt{m_{20} m_{02}} c_2 - \bar{x}_1 \bar{x}_2 c_1^2) / (c_1 \sqrt{m_{20} m_{02}})$$

(4.15)

From these equations it is obvious that the accuracy in the parameters δ and λ is not high, as even fourth order mixed moments are involved. This was also confirmed by calculations of these parameters for some observation series. In many cases values of the coefficients $\{c_k\}$ did not converge, which is an necessary condition. This was usually the case when the estimated values of c_1 and c_2 were approximately equal. In table 4.3 parameters' values from one of the observation points are shown. The result here is good compared to other observation points. For the series from the large Swedish lakes, for instance, convergency was not obtained for less than half of the year.

The inaccuracy in the parameter estimation prevents thus the use of the bivariate gamma distribution eq (4.7). A simplification can, however, be made, so that just two terms in the Laguerre series expansion are used. The parameters c_k can then be chosen so that the sample correlation coefficient is preserved. Compare to equation (6.7).

TABLE 4.3 Monthly parameters for the bivariate gamma distribution at 86-1008 N Forsmöllen

Month	γ	β	c_1	c_2	c_3	c_4	c_5	c_6
1	4.79	3.22						
			0.514	0.298	0.185	0.121	0.082	0.057
2	3.01	5.22						
			0.423	0.374	0.442	0.613	0.943	1.562
3	4.23	3.24						
			0.424	0.178	0.074	0.031	0.013	0.005
4	3.19	4.57						
			0.713	0.685	0.769	0.949	1.249	1.724
5	4.07	2.04						
			0.410	0.063	0.002	0.000	0.000	0.000
6	10.50	0.50						
			0.327	0.047	0.000	0.000	0.000	0.000
7	4.60	1.14						
			0.658	0.226	0.036	0.001	0.000	0.000
8	2.08	2.62						
			0.739	0.267	0.042	0.001	0.000	0.000
9	2.57	2.29						
			0.557	0.113	0.001	0.000	0.000	0.000
10	2.50	3.18						
			0.704	0.189	0.004	0.001	0.000	0.000
11	2.82	3.95						
			0.573	0.125	0.001	0.000	0.000	0.000
12	4.23	3.40						
			0.496	0.335	0.268	0.238	0.228	0.231
1	4.79	3.22						

5. FINITE DIFFERENCE EQUATION FOR THE RUNOFF PROCESS

In this chapter we shall derive a finite difference equation for the monthly runoff process from the multivariate normal distribution functions (eqs 4.10 and 4.12). Such equations are better for simulation purposes than a direct use of the multivariate distribution functions. Other finite difference equations referenced in literature will be also discussed also.

5.1 Scalar process

We shall first consider a scalar process i.e. runoff at one point in a river. This can be described by the multivariate unit normal distribution eq (4.10), where the variables u_i , $i = k, k-1, \dots, k-n$ are normalized modules of runoff for the month i . The exponent in the distribution function can be developed by its cofactors. We now write down eq (4.10) in the following way:

$$f_{n+1}(u_k, u_{k-1}, \dots, u_{k-n}) = \frac{1}{(2\pi)^{n+1} D} \exp \left(-\frac{1}{2D} \sum_{j=0}^n \sum_{g=0}^n (-1)^{j+g} D_{k-j, k-g} u_{k-j} u_{k-g} \right) \quad (5.1)$$

where D is the determinant of the correlation matrix

$$\begin{pmatrix} 1 & \rho_{k, k-1} & \dots & \rho_{k, k-n} \\ \rho_{k-1, k} & 1 & \dots & \rho_{k-1, k-n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-n, k} & \rho_{k-n, k-1} & \dots & 1 \end{pmatrix}$$

and $D_{k-j, k-g}$ are the cofactors to this determinant, derived by excluding the row with index $k-j$ and the column with index $k-g$.

The quadratic form in the exponent can be written as

$$\begin{aligned} & \sum_{j=0}^n \sum_{g=0}^n (-1)^{j+g} D_{k-j, k-g} u_{k-j} u_{k-g} = \\ & = D_{kk} u_k^2 - 2 B u_k + C = \\ & = D_{kk} (u_k - B / D_{kk})^2 + C - B^2 / D_{kk}^2 \end{aligned}$$

where

$$\begin{aligned} B &= D_{k, k-1} u_{k-1} - D_{k, k-2} u_{k-2} + \dots \\ &\dots + (-1)^{n+1} D_{k, k-n} u_{k-n} \end{aligned}$$

and

$$C = \sum_{j=1}^n \sum_{g=1}^n (-1)^{j+g} D_{k-j, k-g} u_{k-j} u_{k-g}$$

The n -dimensional density function can be calculated from the $(n+1)$ -dimensional density function eq 4.1 by formula (2.2).

$$\begin{aligned} f_n(u_{k-1}, u_{k-2}, \dots, u_{k-n}) &= \\ &= \int_{-\infty}^{\infty} f_{n+1}(u_k, u_{k-1}, \dots, u_{k-n}) du_k = \\ &= \int_{-\infty}^{\infty} 1 / ((2\pi)^{n+1} D)^{1/2} \exp(-1/(2D)(D_{kk}(u_k - B/D_{kk})^2 + \\ &+ C - B^2/D_{kk}^2)) \cdot du_k = \\ &= 1 / ((2\pi)^n D_{kk})^{1/2} \exp(-1/(2D)(C - B^2/D_{kk})) \cdot \\ &\cdot 1 / \sqrt{2\pi} \int_{-\infty}^{\infty} \exp(-D_{kk}/(2D)(u_k - B/D_{kk})^2) \\ d(\sqrt{D_{kk}/D} u_k) &= \\ &= 1 / ((2\pi)^n D_{kk})^{1/2} \cdot \exp(-1/2 D (C - B^2/D_{kk})) \end{aligned}$$

The $(n+1)$ -dimensional conditional density function can now be derived by applying eq (2.4):

$$\begin{aligned}
 f_{n+1}(u_k / u_{k-1}, \dots, u_{k-n}) &= \\
 &= \frac{f_{n+1}(u_k, u_{k-1}, \dots, u_{k-n})}{f_n(u_{k-1}, \dots, u_{k-n})} = \\
 &= \frac{1}{(2\pi D / D_{kk})^{1/2}} \exp \left(-\frac{1}{2} \left(\frac{u_k - B / D_{kk}}{\sqrt{D / D_{kk}}} \right)^2 \right) \quad (5.2)
 \end{aligned}$$

This is the formula for the one dimensional normal density function (cf. eq 3.11) with the expected value B / D_{kk} and the variance D / D_{kk} . From (5.2) it follows thus, that

$$\epsilon_k = (u_k - B / D_{kk}) / (D / D_{kk})^{1/2} \quad (5.3)$$

is a normally distributed random variable, with zero mean and the dispersion equal to one. Taking into consideration the full expression for B , we derive the following finite difference equation of the runoff process:

$$u_k = H_{1k} u_{k-1} + H_{2k} u_{k-2} + \dots + H_{nk} u_{k-n} + \Lambda_n \epsilon_k, \quad (5.4)$$

where

$$\begin{aligned}
 H_{jk} &= (-1)^{j+1} D_{k, k-j} / D_{kk} \\
 \Lambda_n &= (D / D_{kk})^{1/2}
 \end{aligned}$$

Λ_n may be expressed by $H_{1k}, H_{2k}, \dots, H_{nk}$. Taking the square of eq (5.4) we get:

$$\begin{aligned}
 u_k^2 &= \sum_{j=1}^n H_{kj}^2 u_{k-j}^2 + 2 \sum_{j=1}^n \sum_{g=1}^n H_{kj} H_{kg} u_{k-j} u_{k-g} + \\
 &+ 2 \Lambda_k \epsilon_n \sum_{j=1}^n H_{kj} u_{k-j} + \Lambda_n^2 \epsilon_k^2 \quad (5.5)
 \end{aligned}$$

$u_k, u_{k-1}, \dots, u_{k-n}, \epsilon_k$ are standardized normal variables and ϵ_k and $u_{k-j}; j \geq 1$ are independent. Taking the expected value of the left and right hand sides of eq (5.5) we get the following relation:

$$1 = \sum_{j=1}^n H_{kj}^2 + 2 \sum_{j=1}^n \sum_{g=j+1}^n H_{kj} H_{kg} \rho_{k-j, k-g} + \Lambda_n^2 \quad (5.6)$$

This gives, thus, an alternative way to calculate the coefficient Λ_n^2 , which can be advantageously used as a test.

It is obvious that eq (5.4) is convenient for the generation of synthetic realisations of the runoff process. Let n values

$u_{k-1, g}, u_{k-2, g}, \dots, u_{k-n, g}$, where g denotes the year, be known. The parameters $H_{1k}, \dots, H_{nk}$ and Λ_n ; $k = 1, \dots, 12$ are calculated from the correlation matrices for each month. ϵ_k , which is a random variable with a unit normal distribution, can be generated from a random number generator on a computer or from a table of random numbers. $u_{k, g}$ can be calculated now from eq (5.4). When $u_{k, g}$ is known, in the same way we can find $u_{k+1, g}$ and so on. To start the simulation it is necessary to know the first n values of $u_{k, g}$ i.e. $u_{1,0}, u_{2,0}, \dots, u_{n,0}$. These values can be chosen quite freely, if we compensate their influence by discarding the first three or four years of simulation. We now have a synthetic series of transformed module values of monthly runoff. The monthly runoff values are derived by solving the set of equations:

$$u_{k, g} = u_k \left(\frac{Q_{k,g}}{Q_k} \right) \quad k = 1, \dots, 12; g = 0, 1, \dots, N \quad (5.7)$$

where N is the number of years simulated. The inverse u_k^{-1} can be given in case of the transformation eqs (3.29) and (3.30), by an explicit expression. For the other transformations given in 3.4.3 eq (5.7) has to be solved by an iterative procedure. In this study the method of Newton-Raphson was used (see, for instance, Fröberg, 1962), which gives fast convergence. Usually from two to four iterations are needed.

Examples of parameters H_{kj} for $n = 1, \dots, 5$, i.e. the order of the Markov process is one to five (cf. eq (2.13)), are shown in Figs 5.1. The coefficients are calculated from the correlation coefficients and therefore show the same kind of seasonal variation as these. This

75-855 Getebro

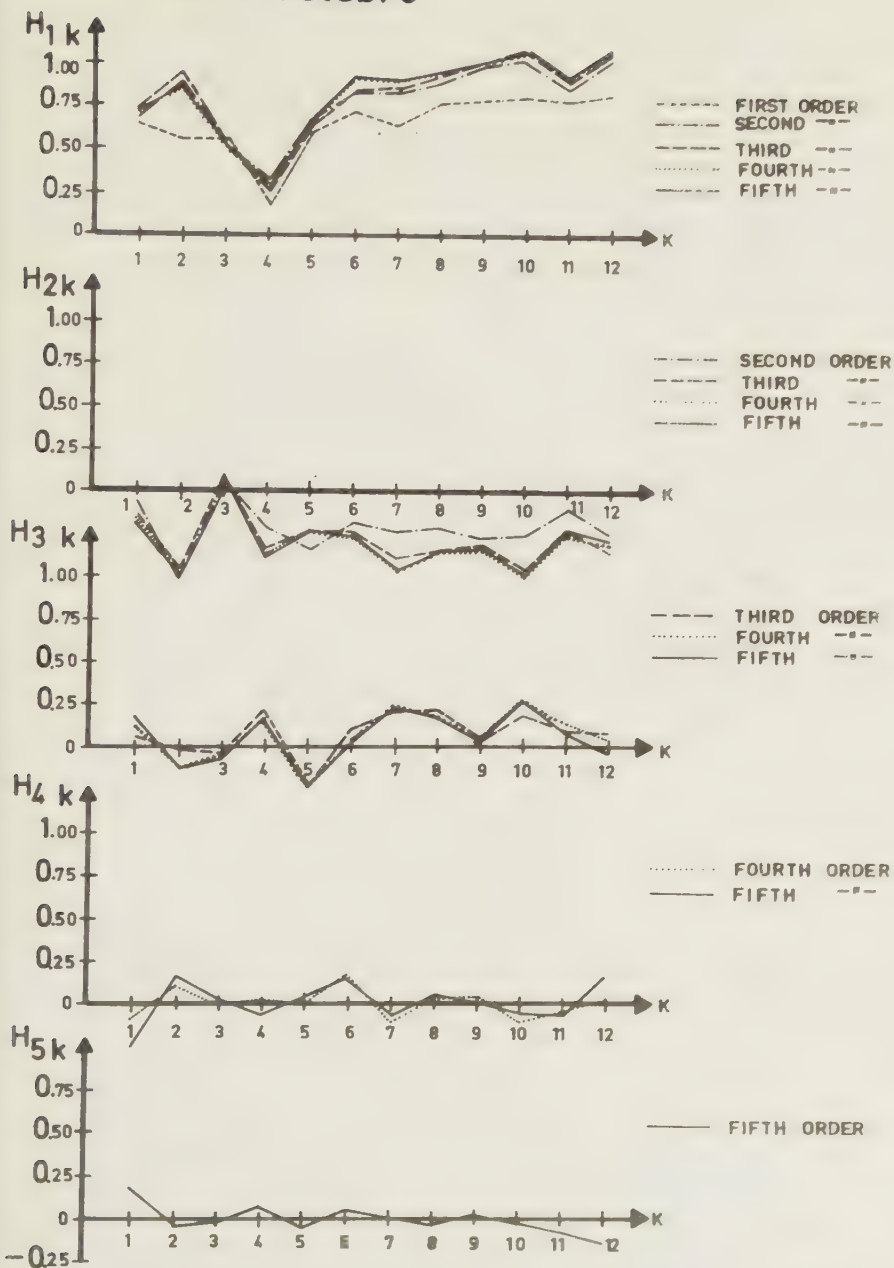


Fig 5.1 Coefficients of first to fifth order stochastic finite difference equation for monthly river runoff for 75-855 Getebro.

is first of all valid for H_{1k} and H_{2k} , while higher order parameters have a more random fluctuation. A significant difference in the parameters H_{1k} can be observed for a first and second order process, respectively. For higher order processes negligible changes are found, compared to the second order parameters. H_{1k} thus seems to be stable for a second order process. In the same way H_{2k} seems stable for a third order process. For still higher orders no significant differences are found. From this we can get some guidance in the choice of order of the process.

5.2 Vector process

We shall now try to generalize eq (5.5) to a vector Markov process. This is the case when we look at runoff at some different points in one and the same river or at points in neighbouring rivers. The derivation that follows is mainly after Kartvelishvili (1972)(personal communication).

We shall use the multivariate unit normal distribution eq (4.12). Ad interim we shall introduce the following new notations:

$$\begin{array}{lll} u_k^1 = \psi_1 & u_{k-1}^1 = \psi_{m+1} \dots & u_{k-n}^1 = \psi_{N-m+1} \\ u_k^2 = \psi_2 & u_{k-1}^2 = \psi_{m+2} \dots & u_{k-n}^2 = \psi_{N-m+2} \\ \vdots & \vdots & \vdots \\ u_k^n = \psi_m & u_{k-1}^n = \psi_{2m} \dots & u_{k-n}^n = \psi_N \end{array}$$

where $N = m(n+1)$, the upper index on u denotes the place and the lower - the number of month. The probability density function of the vector $\{\psi_1, \psi_2, \dots, \psi_N\}$ is expressed by the formula:

$$f_N(\psi_1, \psi_2, \dots, \psi_N) = ((2\pi)^N D)^{-1/2} \exp(-D_u / (2D)) \quad (5.8)$$

D and D_u were defined in connection to eq 4.12. The quadratic form D_u can be written as:

$$D_u = \sum_{j=1}^m \sum_{g=1}^m A_{jg} \psi_j \psi_g + 2 \sum_{g=1}^m B_g \psi_g + C$$

there

$$B_g = A_{g, m+1} \psi_{m+1} + A_{g, m+2} \psi_{m+2} + \dots + A_{g, N} \psi_N$$

$$C = \sum_{j=m+1}^N \sum_{g=m+1}^N A_j \psi_j \psi$$

and A_j the cofactors to the determinant D , by excluding the j -th row and the g -th column. Let us now set $\psi_j = y_j + \alpha_j$ and D_u can be expressed:

$$D_u = \sum_{j=1}^m \sum_{g=1}^m A_{jg} y_j y_g + \sum_{j=1}^m \sum_{g=1}^m A_{jg} (\alpha_j y_g + \alpha_g y_j) +$$

$$+ 2 \sum_{g=1}^m B_g y_g + \sum_{j=1}^m \sum_{g=1}^m A_{jg} \alpha_j \alpha_g + 2 \sum_{g=1}^m B_g \alpha_g + C \quad (5.9)$$

The α_j , $j = 1, 2, \dots, m$ can be chosen so that the following relation is satisfied:

$$\sum_{j=1}^m \sum_{g=1}^m A_{jg} (\alpha_j y_g + \alpha_g y_j) + 2 \sum_{g=1}^m B_g y_g = 0$$

The α_j , $j = 1, 2, \dots, m$ are then the solutions to the equation system below:

$$A_{11} \alpha_1 + A_{12} \alpha_2 + \dots + A_{1m} \alpha_m = -B_1$$

$$A_{21} \alpha_1 + A_{22} \alpha_2 + \dots + A_{2m} \alpha_m = -B_2$$

...

$$A_{m1} \alpha_1 + A_{m2} \alpha_2 + \dots + A_{mm} \alpha_m = -B_m$$

Applying Cramers rule, the solutions can be written as:

$$\alpha_j = H_{j, m+1} \psi_{m+1} + H_{j, m+2} \psi_{m+2} + \dots$$

$$+ H_{j, N} \psi_N; j = 1, 2, \dots, m \quad (5.10)$$

where

$$H_{jg} = 1 / E \cdot (-1)^j F_i; A_{1g} + (-1)^{j+1} A_{2g} + \dots$$

$$+ (-1)^{j+m-1} A_{mg} F_{i, g}$$

$$E = \begin{vmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & & A_{mm} \end{vmatrix}$$

and E_{jg} is the minor to the determinant E derived by excluding the j -th row and the g -th column.

We can now write down the expression for D_u in the following way:

$$\begin{aligned} D_u = & \sum_{j=1}^m \sum_{g=1}^m A_{jg} y_j y_g + \sum_{j=1}^m \sum_{g=1}^m A_{jg} \alpha_j \alpha_g + \dots \\ & \dots + 2 \sum_{j=1}^m B_j \alpha_j \end{aligned} \quad (5.11)$$

Under the condition that the α_j , $j = 1, 2, \dots, m$ satisfy the formula (5.10) the probability density function for the vector $\{\psi_1, \psi_2, \dots, \psi_N\}$ will have the expression:

$$\begin{aligned} f_N(\psi_1, \psi_2, \dots, \psi_N) = & ((2\pi)^N D)^{-1/2} \cdot \\ & \cdot \exp \left\{ -1/(2D) \left(\sum_{j=1}^m \sum_{g=1}^m A_{jg} \alpha_j \alpha_g + 2 \sum_{g=1}^m B_g \alpha_g + C \right) \right\} \cdot \\ & \cdot \exp \left\{ -1/(2D) \sum_{j=1}^m \sum_{g=1}^m A_{jg} y_j y_g \right\} \end{aligned} \quad (5.12)$$

We can calculate the probability density function of the vector $\{\psi_{m+1}, \psi_{m+2}, \dots, \psi_N\}$ from the distribution (5.12) by applying eq (2.2):

$$\begin{aligned} f_{N-m}(\psi_{m+1}, \psi_{m+2}, \dots, \psi_N) = & \\ = & \overbrace{\int \dots \int}^{\infty} f_N(\psi_1, \psi_2, \dots, \psi_m, \psi_{m+1}, \dots, \psi_N) \\ & d\psi_1 d\psi_2 \dots d\psi_m = (D^{m-1} / (2\pi)^{N-m})^{1/2} \cdot \\ & \cdot \exp \left\{ -1/2 D \left(\sum_{j=1}^m \sum_{g=1}^m A_{jg} \alpha_j \alpha_g + 2 \sum_{g=1}^m B_g \alpha_g + C \right) \right\} \end{aligned} \quad (5.13)$$

Using a generalization of eq (2.4) we can now find an expression for the conditional distribution function from eqs (5.12) and (5.13):

$$\begin{aligned}
 f(\psi_1, \dots, \psi_m | \psi_{m+1}, \dots, \psi_N) &= \\
 &= f_N(\psi_1, \dots, \psi_m, \psi_{m+1}, \dots, \psi_N) / \\
 f_{N-m}(\psi_{m+1}, \dots, \psi_N) &= ((2\pi)^m D)^{-1/2} \\
 \exp \left\{ -1/(2D) \sum_{j=1}^m \sum_{g=1}^m A_{jg} y_j y_g \right\} & \quad (5.14)
 \end{aligned}$$

The absolute positive quadratic form in the exponent can be transposed to a sum of squares, i.e. transposed to the form:

$$\sum_{j=1}^m \sum_{g=1}^m A_{jg} y_j y_g = \sum_{j=1}^m \left(\sum_{g=1}^m \mu_{jg} y_j \right)^2 = \sum_{j=1}^m \xi_j^2 \quad (5.15)$$

We can develop the expression for the ξ_j ; $j = 1, \dots, m$ in the following way:

$$\begin{aligned}
 \xi_j &= \sum_{g=1}^m \mu_{jg} y_g = \sum_{g=1}^m \mu_{jg} (\psi_j - \alpha_j) = \\
 &= \mu_{j1} \psi_1 + \dots + \mu_{jm} \psi_m - \nu_{j, m+1} \psi_{m+1} - \dots \\
 &\quad \dots - \nu_{j, N} \psi_N \quad (5.16)
 \end{aligned}$$

where

$$\nu_{ji} = \mu_{j1} H_{1i} + \mu_{j2} H_{2i} + \dots + \mu_{jm} H_{mi}$$

The parameters μ_{jg} in the formulas above are dependent on the parameters A_{jg} only. This dependence is not unique (see, for instance, Smirnov, 1958, p 120) and the parameters μ_{jg} can be chosen in several ways. For any of them we can write down the density function (4.14) as follows:

$$\begin{aligned}
 f(\psi_1, \dots, \psi_m | \psi_{m+1}, \dots, \psi_N) &= \\
 &= ((2\pi)^m D)^{-1/2} \exp \left(-1/2 \sum_{j=1}^m \xi_j^2 / D \right) \quad (5.17)
 \end{aligned}$$

Substituting

$$\xi_j / \sqrt{D} = \epsilon_j \quad (5.18)$$

to eq (5.17) we get a product of m unit normal density functions, i.e. ϵ_j is an independent random variable normally distributed with zero mean and variance equal to one. In an open form eq (5.18) can be written as:

$$\begin{aligned} & \mu_{j1} \Psi_1 + \mu_{j2} \Psi_2 + \dots + \mu_{jm} \Psi_m = \\ & = \nu_{j, m+1} \Psi_{m+1} + \nu_{j, m+2} \Psi_{m+2} + \dots + \nu_{j, N} \Psi_N + \\ & + \sqrt{D} \epsilon_j \end{aligned} \quad (5.18')$$

We have m expressions (5.18') with $j = 1, \dots, m$ and we can write them in one single matrix equation. If we return now to the original variables u_i^s ; $i = k, k-1, \dots, k-n$; $s = 1, 2, \dots, m$ (cf. page 156), this equation will have the expression:

$$\vec{u}_k = H_1 \vec{u}_{k-1} + H_2 \vec{u}_{k-2} + \dots + H_n \vec{u}_{k-n} + \nu^{-1} \sqrt{D} \vec{\epsilon} \quad (5.19)$$

here

$$\mu = \begin{pmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2m} \\ \vdots & \vdots & & \vdots \\ \mu_{m1} & \mu_{m2} & \dots & \mu_{mm} \end{pmatrix}$$

$$H_s = \begin{pmatrix} H_{1, s, m+1} & H_{1, s, m+2} & \dots & H_{1, (s+1), m} \\ H_{2, s, m+1} & H_{2, s, m+2} & \dots & H_{2, (s+1), m} \\ \vdots & \vdots & & \vdots \\ H_{m, s, m+1} & H_{m, s, m+2} & \dots & H_{m, (s+1), m} \end{pmatrix}$$

and

$$\vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix}$$

Let us now turn to matrix μ . It is reasonable to choose it orthogonal, i.e. such that

$$\mu_{jg} = \sqrt{\lambda_j} \beta_{jg}$$

where λ_j ; $j = 1, \dots, m$ are the solutions to the secular equation:

$$\begin{vmatrix} A_{11} - \lambda & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} - \lambda & \dots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} - \lambda \end{vmatrix} = 0 \quad (5.20)$$

and (β_{jg}) the corresponding unitar matrix which has the property:

$$(\beta_{jg}) \times (\beta_{gj}) = I \quad (5.21)$$

The matrix μ is thus written:

$$\mu = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_m} \end{pmatrix} \times (\beta_{jg})$$

The matrix (A_{jg}) is symmetric and therefore all the solutions to equation (3.20) $\lambda_1, \dots, \lambda_m$ are real, the matrix (β_{jg}) is also symmetric and its inverse equals the transposed matrix (β_{gj}) .

The inverse to the matrix μ is thus calculated as:

$$\mu^{-1} = (\epsilon_{gj}) \times \begin{pmatrix} \sqrt{\lambda_1^{-1}} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2^{-1}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_m^{-1}} \end{pmatrix}$$

The j -th component η_j of the random vector $\mu^{-1} \vec{\epsilon}$ has the expression:

$$\eta_j = \sqrt{\lambda_1^{-1}} \epsilon_{1j} \epsilon_1 + \sqrt{\lambda_2^{-1}} \epsilon_{2j} \epsilon_2 + \dots + \sqrt{\lambda_m^{-1}} \epsilon_{mj} \epsilon_m \quad (5.23)$$

The η_j ; $j = 1, 2, \dots, m$ have the following two properties:

- 1) the expected value $E(\eta_j)$ is equal to zero as the random variables $\epsilon_1, \epsilon_2, \dots, \epsilon_m$ have the expected value equal to zero.
- 2) η_j ; $j = 1, 2, \dots, m$ are normally distributed as they are linear functions of the normally distributed variables $\epsilon_1, \epsilon_2, \dots, \epsilon_m$

We finally write down equation (5.19):

$$\vec{u}_k = H_1 \vec{u}_{k-1} + H_2 \vec{u}_{k-2} + \dots + H_n \vec{u}_{k-n} + \Lambda \vec{\epsilon} \quad (5.24)$$

where $\Lambda = \sqrt{D} (\sqrt{\lambda_j^{-1}}) \times (\epsilon_{gj})$. For the j -th component u_k^j of the vector $\vec{u}_k$ the expression will be:

$$\begin{aligned} u_k^j &= H_{j, m+1} u_{k-1}^1 + H_{j, m+2} u_{k-2}^1 + \dots + H_{j, 2m} u_{k-m}^1 + \dots \\ &\dots + H_{j, m \cdot n+1} u_{k-n}^{nn} + \dots + H_{j, m(n+1)} u_{k-n}^m + \\ &+ \sqrt{D} (\sqrt{\lambda_1^{-1}} \epsilon_{1j} \epsilon_1 + \dots + \sqrt{\lambda_m^{-1}} \epsilon_{mj} \epsilon_m) \end{aligned} \quad (5.25)$$

This is the generalization of eq (5.4) for a vector process. It is necessary to remember that the matrices μ and H_s ; $s = 1, \dots, m$ are different for different months k .

5.3 The order of the Markov process

It is a complicated problem to determine the order of the Markov process. The determination is based on the information we can get from the correlogram. By this the correlogram as a whole will be considered. We have already noted that from a theoretical point of view the correlation coefficients are subjected to errors and biases. These are mainly caused by too short observation series. The construction of the model is, furthermore, based on some assumptions about the studied process. We have, for instance, assumed that the process is Markovian, that it is linear and that it is normal. Errors arise also because of the fact that the natural process does not possess these features. The criteria, that we can set up for the choice of the order of the process are also based on these assumptions.

Any true order does not, of course, exist. We can only talk about the best order based on some certain criterion. This criterion can be defined in many different ways. Applying different criteria we shall get contradictory results. Below we shall discuss in more detail some of criteria to determine the order of the Markov process. We shall also try to find some regional patterns of this order.

As we are dealing with a nonstationary process the order can be different for different months. The order is then determined by the maximum order over the twelve months.

Below we have used the correlation coefficient of the original sample to calculate the coefficients of eqs (5.4) and (5.24). We assume that the results and general conclusions about the order shall be the same if correlation coefficients calculated in other ways are used.

5.3.1 Test of normality of the residual

The hypothesis is that the residuals calculated from the data of N observations are unit normally distributed. We shall accept a χ^2 -test (see, for instance, Cramer, 1948, p 416). The space of the normal variable is divided into eight parts with equal probability masses, $p_i = 1/8$. The test quantity will be:

$$\chi^2 = \sum_{i=1}^8 \frac{(v_i - N \cdot 1/8)^2}{N \cdot 1/8}$$

where v_i is the number of observations in the i -th interval. χ^2 is distributed in a χ^2 -distribution with 7 degrees of freedom, if N is large.

Test quantities were calculated for residuals of first to fifth order processes for the 38 rivers. For those 456 months the hypothesis, that the first order residual was unit normal, was rejected on 5 % level in 28 (6,5 %) cases. As the order increases there is a tendency that the number of rejections increases (table 5.1). For 1 % and 0,1 % significance the results are the same. No regional regularities could be traced.

TABLE 5.1 Number of rejections of hypothesis that the residuals are unit normally distributed.

Order of model	Level of significance	5 %	1 %	0,1 %
1		28 (6,5 %)	3 (0,7 %)	2 (0,5 %)
2		30 (6,9 %)	9 (2,1 %)	2 (0,5 %)
3		27 (6,2 %)	10 (2,3 %)	2 (0,5 %)
4		42 (9,7 %)	10 (2,3 %)	3 (0,7 %)
5		34 (7,9 %)	16 (3,7 %)	3 (0,7 %)

5.3.2 Test of independence of the residual

The adequacy of the hypothesis of independence is not easily tested in the case of autoregression. Serial correlations are generated in the observed residuals, which makes it impossible to test the original model for independence of residuals merely by testing the observed residuals for independence (Kendall, 1973). In case of regression relationship Durbin and Watson (1971) have suggested a

more sophisticated test. This test can not be applied to autoregression series (Durbin, 1970 Kendall, 1973). In hydrology test of independence of the residual have been used to determine the order (Yevjevich, 1972a). The order is increased stepwise until the residual is considered to be independent by some test. There exist several tests for serial independence, a review of them is, for instance, given in Wallis and Matalas (1971). In hydrologic studies, one of the most commonly used tests is the Anderson (1942) test.

Residual correlation coefficients for models of order one to five were calculated for the 38 observation series. When applying the Anderson test the hypothesis of independence of residuals could be rejected for first order models and in most cases for second order models. No regional patterns could be traced.

5.3.3 Test on variance of the residual Λ^2

The relation between the residual standard deviation Λ and the order n of the Markov process will have the principal form shown by the example in Table 5.2. With the growth of the order the standard deviation will decrease. In general, the difference between different orders is relatively large for low orders and reduces for higher orders. One could expect that the standard deviation strives to an asymptotic minimal value. It is clear from the example shown that a considerable reduction in the standard deviation of the residual is gained when turning from the independent to first order model. For some months another noticeable decrease is observed between the first and the second order models. For higher order models the gain of information by a decrease in the residual standard deviation is usually negligible.

Taking into consideration the degrees of freedom that are lost by the estimated parameters and growth with the order, the relation $\Lambda(n)$ will have a minimum. The minimum is more or less well defined. Jenkins and Watts (1968), suggesting this way of determination of the order of an autoregressive process, advise to use it together with some other method.

TABLE 5.2 Residual standard deviation for different months and for different orders of Markov process for 28-444 Hällbacken

Month	Order					
	0	1	2	3	4	5
1	1.0	0.438	0.436	0.432	0.432	0.432
2	1.0	0.308	0.285	0.285	0.281	0.279
3	1.0	0.331	0.322	0.318	0.318	0.318
4	1.0	0.419	0.400	0.400	0.399	0.395
5	1.0	0.950	0.938	0.927	0.922	0.920
6	1.0	0.997	0.929	0.919	0.918	0.904
7	1.0	0.976	0.895	0.877	0.873	0.872
8	1.0	0.957	0.956	0.919	0.919	0.917
9	1.0	0.906	0.869	0.829	0.822	0.814
10	1.0	0.918	0.900	0.861	0.853	0.829
11	1.0	0.725	0.720	0.714	0.711	0.704
12	1.0	0.507	0.497	0.496	0.474	0.474

Let us consider equation 5.4 as an ordinary regression equation. Actually we have autocorrelation both in the u 's and the residuals. The correlation for one month, from year to year is, however, very small and this is usually also valid for the residual autocorrelation (cf 4.3.2). The effect of autocorrelation in the variables and residuals reduces the reliability of the estimated parameters (Wold, 1950 and Lyttkens, 1963). We shall use a F-test (see Brownlee, 1967) to see if any significant improvement can be obtained by using a model of order $n+1$ instead of order n . A similar test can be applied to determine the order of more general stochastic finite difference equations (Åström, Bohlin and Wensmark, 1965). The null hypothesis is that we can make the fit of the regression equation by n parameters and the alternative hypothesis is that one more parameter is necessary. The test quantity is given by the

following expression:

$$F^n = \frac{\Lambda_n^2 - \Lambda_{n+1}^2}{\Lambda_{n+1}^2} \frac{(N - 1 - (n+1))}{((n+1) - 1)}$$

which under the null hypothesis is distributed as $F(1, N - 1 - (n+1))$. Test quantities F_k^n , where k indicates individual months, were calculated for $k = 1, \dots, 12$ and for different n 's for 38 observation series. For each of the series the dependence $\max |F_k^n|$ on n was obtained and the test was performed on the values got. Not in a single case the hypothesis, that a first order model fits, could be rejected on a 99,9 % level. For 5 rivers the hypothesis was rejected on 99 % - level and for 20 rivers on 95 % - level. The hypothesis that the second order model fits could not be rejected in any case on the three significant levels 99,9 %, 99 % and 95 %, except for the outflow from the great lakes, where even the hypothesis, that the third order model fits, could be rejected on the 95 % - level.

5.3.4 Test on the regression coefficients H_{nk}

The second regression coefficient H_{2k} in a second order Markov process will have the following expression:

$$H_{2k} = \frac{\rho_{k, k-2} - \rho_{k, k-1} \rho_{k-1, k-2}}{1 - \rho_{k-1, k-2}^2}$$

If the river runoff actually is a first order Markov process, then $\rho_{k, k-2} = \rho_{k, k-1} \rho_{k-1, k-2}$ and H_{2k} is equal to zero. The result can be generalized to higher orders. To test the order n we set up the hypothesis that $H_{n,k}$ is equal to zero. Under the assumption of normality, we can use the test quantity:

$$t_{nk} = \sqrt{N - n} \frac{D_{nk}}{D_{nk}} H_{nk}$$

which is distributed in Student's distribution with $N - n$ degrees of freedom (Cramer, 1948). As our aim is to find the proper order for all months of the year, we make the test on the quantities $\max_k |t_{nk}|$ for different n .

Application of this test gives a rather puzzling-picture and a proper order is difficult to determine. For almost every observation series there is one or more months for which the hypothesis that H_{nk} is zero can be rejected on 95 % or 99 % level, for even very high orders. This is specially the case when the first and second order correlations are relatively low i.e. for the spring flow period. On 99,9 % level the hypothesis can be rejected for first order models and in most of the cases also second order models.

5.3.5 Other tests

A test similar to the test of H_{nk} is the use of the partial correlation coefficient, which is expressed by:

$$\rho_{k, k-n; k-1, \dots, k-n+1} = \frac{D_{k, k-n}^{nk}}{D_{kk}^{nk} D_{k-n, k-n}^{nk}}$$

For a second order process we get:

$$\rho_{k, k-2; k-1} = \frac{\rho_{k, k-2} - \rho_{k, k-1} \rho_{k-1, k-2}}{(1 - \rho_{k, k-1}^2) (1 - \rho_{k-1, k-2}^2)}$$

One can see that in the same way as H_{2k} , the partial correlation coefficient $\rho_{k, k-2, k-1}$ vanishes if we are considering a first order process in reality. While the correlogram can extend to infinity, the partial correlogram of the right order vanishes. We can thus test the hypothesis that the partial correlation is equal to zero. Such test was performed only on one river for some months and as it gave the same result of the test as H_{nk} , it was not extended further.

A more sophisticated test of the order of stationary autoregressive processes is suggested by Quenouille (1948). He uses the property of the partial correlations (Kendall, 1973), but avoids the explicit calculations of the partials. The technique is rather laborious but it nevertheless has been applied in hydrology (Quimpo, 1967).

5.3.6 Conclusions about the order

From the performed tests of the order, it seems to be proper to choose the second order model as the model of best fit. This can be said to be valid for all of the rivers studied except the great lakes outflows, and no regional patterns can be traced. The different tests give to some extent contradictory results. In such cases the test of the regression coefficients has been considered to be erroneous. The test of normality of the residual does not actually point out that a certain order is better than the others, except for some tendency to prefer lower orders. The test of independence of the residual and the F-test do coincide in most cases.

It is, of course, a still more complicated problem to determine the order of the vector model. In principle, the same methods as those applied to the scalar model could be possible. For instance, the standard deviation of the residual for each vector component as a function of the order shows the same decreasing dependency as for the scalar model, which is illustrated by three examples in Table 5.3. Complications can arise when for all the components totally we observe a decrease, but individual components can show an increase.

To test on the regression matrix coefficients Kartvelishvili (1972) (personal communication) suggests to use the norm of the matrix $|H_{nk}|$. In the same way as for the vector model we should determine the dependence $\max_k |H_{nk}|$ and then test those values against some fixed value δ . This dependence is not monotonously decreasing but shows the same kind of instabilities as the coefficients of the scalar process, especially for larger lags. The individual components of the matrices show the same dependences.

The number of parameters increases very fast when the order grows. If we consider also the behaviour of the scalar model a general conclusion is to use a first order vector model.

TABLE 5.3 Examples of residual standard deviation of vector models for different orders of Markov process for January.

Vector component	Order			
	1	2	3	4
Observation points 1-959, 1-3 and 1-589				
1	0.579	0.615		
2	0.590	0.393		
3	0.460	0.367		
Observation points 28-444 and 28-56				
1	0.435	0.405	0.336	
2	0.499	0.507	0.397	
Observation points 98-1048 and 98-219				
1	0.601	0.444	0.398	0.384
2	0.853	0.625	0.542	0.521

5.4 Generation scheme for the bivariate gamma distribution

The bivariate gamma distribution cannot be expressed by a finite difference equation like the one derived for the normal distribution. We can give a regression equation for each component in the Laguerre expansions (Lancaster, 1958). The residual distribution, however, is not known. To generate correlated gamma distributed series, we shall use the conditional gamma distribution directly i.e.:

$$F(x_2 | x_1) = f(x_2; \gamma_2, \beta_2) \sum_{k=0}^{\infty} c_k L_k^{\gamma_1-1}(x_1 \beta_1) L_k^{\gamma_2-1}(x_2 \beta_2) \quad (5.26)$$

If x_1 is known, x_2 can be generated from a rectangularly distributed number ϵ by solving the equation $F(x_2 | x_1) = \epsilon$. This is a very time

consuming process that makes it difficult to apply the bivariate gamma distribution for the generation of synthetic runoff series.

5.5 The use of the stochastic finite difference equation for river runoff

In the most general case the stochastic finite difference equation for a Markov process can be written in the form:

$$\phi(\vec{x}_k, \vec{x}_{k-1}, \dots, \vec{x}_{k-n}, \vec{\epsilon}_k) = 0 \quad (5.27)$$

where ϵ_k^x is a random vector and ϕ is some deterministic vector function. Let us consider the most simple case when the function ϕ is linear. Eq (5.27) then turns to the following equation with matrix coefficients:

$$\vec{x}_k = L_{k-1} \vec{x}_{k-1} + \dots + L_{k-n} \vec{x}_{k-n} + M_k \vec{\epsilon}_k \quad (5.28)$$

Models of this principal form are frequently referenced in literature as the way to simulate the river runoff process. The order n of the Markov process in most cases is not determined but only postulated. Further, in principle, there is no need for some hypothesis about the distribution of the residual vector. The assumption of normality and independence of the vector $\vec{\epsilon}_k$ is, however, fundamental in many aspects. As we have seen earlier, it is usually only for this case that a theoretical derivation of biases and errors in the parameters can be done and the sample distributions can be determined. We have also seen, that when scwness is present, errors and biases are usually larger. This is also valid, when we have dependence in the residuals. The application of the maximum likelihood method usually also demands the assumption of normality.

The dilemma is that runoff data cannot be considered to be distributed normally. The assumption of linearity is probably also a rough approximation. We have faced the problem above and suggest the following model for monthly river runoff:

$$\vec{x}_k = \vec{u}_k^{-1} \{ H_{1k} \vec{u}_{k-1} (\vec{x}_{k-1}) + \dots + H_{nk} \vec{u}_{k-n} (\vec{x}_{k-n}) + A \epsilon \}, \quad k = 1, \dots, 12 \quad (5.29)$$

where $\vec{u}_k^{-1}$ is the inverse to the function $\vec{u}_k$. This model is in general non-normal and non-linear. The extent of its non-normality and non-linearity depends on the transformation functions $\vec{u}_k$; $k = 1, \dots, 12$. Linear Markov regression models for the simulation of river runoff have been used by several authors. Svanidze (1961) has suggested the following model for annual flow:

$$Q_i = \bar{Q} + r (Q_{i-1} - \bar{Q}) + f_i (\xi_i, C_{v,i}^{\text{cond}}) \sqrt{1-r^2} \quad (5.30)$$

where ξ_i is a rectangular distributed random number, $C_{v,i}^{\text{cond}}$ - the conditional variance coefficient, given by:

$$C_{v,i}^{\text{cond}} = \frac{2 s \sqrt{1-r^2}}{\bar{Q} + r (Q_{i-1} - \bar{Q})}$$

and $f_i (\xi_i, C_{v,i}^{\text{cond}})$ the ordinate of the two parameter Γ -distribution eq (3.34), corresponding to the given random number and conditional coefficient of variance. Fiering (1961) uses a first order Markov model derived from the bivariate normal distribution. Svanidze (1964) gives a generalized expression of the model eq (5.30) for an arbitrary order of the process. An expression of analogous type is used by Yevjevich (1964). A general theory of this class of models can, for instance, be found in Box and Jenkins (1970) and Kendall (1973). These models can be described by an expression of the form:

$$\left(\sum_{j=1}^n \alpha_j B^j \right) x_t = \epsilon_t \quad (5.31)$$

where ϵ_t is the random component, α_j autoregression coefficients, and B - the backward shift operator $B^n x_t = x_{t-n}$. In general, this process is not stationary and to ensure this, the roots of the eq $\sum_{j=1}^n \alpha_j z^j = 0$, should lie inside the unit circle.

It has been questioned if a Markov model of the form eq (5.27) describes the true nature of annual hydrologic time series. The discussion in hydrologic literature has been concentrated, first of all, upon the

asymptotic behaviour of the range of cumulative departures from the mean for a given sequence of runoff of N years. The essence of this discussion can be found in Salas-La Cruz, J D and Boes, (1974). The so called Hurst coefficient h characterizes the asymptotic behaviour of the expected value of the rescaled range. For a Markov process it has the theoretical value 0.5. Hurst was, however, the first to notice that series, observed in nature, have values of h greater than 0.5. There exists, however, a transient region for Markov processes, where $h > 0.5$ and with the growth of N h strives to 0.5. This convergence can be very slow even for a first order model. Alongside with this explanation of the Hurst phenomena, there exist two more, that have been cited in literature (O'Connell, 1971), namely the effect of marginal distributions and the autocorrelation effect. Simulation studies by Matalas and Huzzzen (1967) and by Mandelbrot and Wallis (1969d) disregard the effect named first, as a possible reason to the Hurst phenomena. Several new models have been introduced, which have an autocorrelation structure that preserves the Hurst phenomena. The first of these models was the so-called fractional Gaussian noise (Mandelbrot and Wallis 1968, 1969a, b, c, d). Later the "broken line" process (Meija 1971, Rodriguez - Iturbe et al. 1972) and the Arima (1, 0, 1) - process (O'Connell, 1971) have been used for this purpose. The specific problems for the application of all these models are those of parameter estimation. Streamflow series are too short to show asymptotic properties.

To apply the model eq (5.30) to monthly flow the instationarity, due to the annual periodicity of river runoff, should be removed. This can be done by standardizing the variables, which will result in a weekly stationary process. This is done in the simplest way by the direct use of the calculated monthly means and standard deviations. It can also be done by fitting Fourier series to the monthly periodic movements of these parameters. This latter approach can be a way to reduce the number of parameters in a monthly simulation model and it was used by Roesner and Yevjevich (1966). A generalization of the model eq (5.30), for monthly flows is given by Reznikovskij (1969). We shall further mention the model suggested by Beard (1965), which is based on regression with a log Pearson type III transformation.

An alternative way to simulate monthly flow is to begin with a simulated annual series and then for each particular year generate the monthly flows. Such methods are the method of fragments suggested by Svanidze (1964) and the use of a so called disaggregation process (Valencia and Schaake, 1973). This is a way to preserve relevant properties of annual flow, that monthly models cannot do (cf 6.2).

Several models for multivariate simulation are also cited in literature. Markovian models are suggested by Fiering (1964), Matalas (1967), Young and Pisano (1968), Reznikovskij (1969) and Crosby and Maddock (1970). With these models it is possible to reproduce the first three moments and the lag-one serial correlation coefficients of each of the series as well as the lag-zero and lag-one cross-correlation coefficients among them. For the non-Markovian case models have been worked out by Matalas and Wallis (1971) and Meja (1971). These models reproduce the value of n , the three first moments, the lag-zero cross-correlation and besides for the first model - the first serial correlation and for the second model - the second spectral moment.

6. PROPERTIES OF THE MODELS

The right choice of a mathematical model to describe the random characteristics of river runoff is a difficult task. In our case we should choose the marginal distribution and the order of the Markov process to define our model. We have assumed that in this way the nature of the runoff process can be described. This is, of course, only an approximation and we cannot account that our model actually can give a total description of the complex nature of runoff. We can hope that our model reflects those properties of the runoff process that are of importance for our specific problem. Various planning situations set different demands to the model.

To describe the properties of the runoff series we usually use statistical parameters, like empirical moments. Our model will preserve the empirical moments only in the case, when they have been used directly in order to determine the parameters of the model. Otherwise we can be sure only that our model will preserve its own parameter values. We have already noted that in many cases the moments can be erroneous, for instance, when the data is skewed. In these cases one can expect that other methods of estimation give better parameters.

Other statistical properties of the runoff process, like crossing properties, extreme values and the asymptotic behaviour of the rescaled range will be preserved, depending on the extent, to which the basic assumptions about the type of the marginal and multivariate distribution and the Markovian character of the process are satisfied. If our assumptions are good, the parameters in the model will really possess the information about the properties, mentioned above. To assure that the model will have a certain property, we can introduce in it parameters that directly characterize this property. Examples are the broken line process and the fractional noise model, where specific parameters - the second order moment of the correlation function and the Hurst coefficient, respectively - are involved in the model structure to give the models the necessary crossing properties and behaviour of the rescaled range. The physical interpretation of such parameters is, however, difficult.

Of large importance is, of course, to see the influence of different models on the results of water management calculations. Some studies of this problem are referenced in literature. Most of them deal with reservoir design for annual flow. Ratkovitj (1968a, 1968b, 1969, 1970) tests five different models based on the gamma distribution, among them the models eqs (4.6), (4.9) and (5.29). The study is mainly concentrated on the extent to which the models show the same grouping tendencies of high-water and low-water years and their influence on water regulation calculations, for instance, the security of firm deliverance of water from a reservoir. The conclusion is, that the model eq (4.9) satisfies all the criteria set up, while the other considered models underestimate reservoir volumes and give different grouping tendencies of high-water and low-water years than those, given by the observed series of annual flow. Burges and Linsely (1971) study factors influencing reservoir storage and stress the importance of good estimates of the variance and the correlation of runoff. They also draw attention to the fact that different results are achieved when using monthly or annual runoff in a model and recommend monthly values. Askew et al (1971) test to which extent ten different models can reproduce the same critical periods as those got from real runoff records. Critical period is defined as a period of time in which the hydrologic record would have been most critical with respect to the demands of the water management system. Their conclusions are that the fractional noise model simulates best characteristics of the critical period as maximum permissible extraction rate from a hypothetical reservoir, length of the critical period and the accumulated deficiency relative to the mean flow. If the costs for the use of this model are too great then a first order log Markov model or the model, suggested by Beard (1965) are recommended. Wallis and Matalas (1972) compare a first order Markov model and a fractional noise model for the calculation of the minimum capacity of a reservoir for certain demands. For demands less than 0.8 of the mean annual runoff the Markovian model is satisfactory, while for higher demands the fractional noise model should be used.

Dealing with design of large reservoirs, the most important property of a model is the correlation structure. The marginal distributions as well as the distribution of the residual have very small influence (Matalas and Huzzen, 1967, Rodriquez-Iturbe et al, 1971). Klemes

(1972) has, however, pointed out in a comment to the article by Rodriguez-Iturbe that for low levels of development the storage capacity depends significantly on the input distribution type for all risks of failure down to values of $< 1\%$. Kottegoda (1974) has shown the significant influence of skewness on the crossing properties.

The model presented here is determined by its marginal distribution functions and the order of the Markov process. Below we shall test to which extent these two factors influence the ability of the model to preserve the moments and the crossing properties of the observed series. An interesting question, that we have not met in literature, is how a model actually preserves the properties of the annual process for periods less than one year. Some comments on this will also be given. As a base for the discussion simulated series from the observation stations 96-1008 N Forsmölla, 1-958 Kallio and 53-121 Fäggeby were used. Ten simulations with the length, equal to that of the observed series, were made with models of different orders and with different distributions. This is, of course, not satisfactory for a more thorough analysis of the properties of the different models.

6.1 Moments

The method of moments can be used to determine the parameters in our model. In this case we can be sure that our model will preserve the moments of the observed series. When other methods are used for parameter estimation, for instance, the maximum likelihood method we can not be certain that the moments are preserved, but only the parameters of the model. In some studies referenced in literature the condition of how well the moments are preserved is used as quality criteria of a certain model (Beard 1965, Burges 1972, McMahon et al 1972). To preserve the observed moments is, of course, an important property of a model. The question is, however, which parameters describe the process best in a certain case. We believe that in cases with high coefficient of variation and skewness the information of the process is preserved better, when using other methods of estimation, although this implies that the moments of the process are not preserved. We shall further

note that it is usually in those cases, that large differences are observed.

To calculate the moments of a certain model the Monte-Carlo method can be used (see, for instance, McMahon et al, 1972). Often it is not necessary, as we can calculate the moments theoretically. In case of the transformations of the unit normal distribution eqs 3.31 - 33 and 3.22, we need simulated series from which the moments can be calculated, but for all other distribution functions, presented here, theoretical expressions can be given. We shall note also that applying the Monte-Carlo method, it is necessary to use very long series to get good estimates of the moments, a thousand or several thousands of year the longer - the higher order of the moments.

Applying the normal distribution eq (3.10) to describe monthly runoff we shall preserve the mean values, the variance and covariance of the observed series, as the theoretical first and second order moments of the marginal and multivariate normal distribution are identical to them.

The log-normal distribution will not necessarily preserve the observed moments exactly, when any of the methods of parameter estimation presented in (3.4.5) is used. The theoretical expected value, α_1 , variance μ_2 and skewness of the marginal two parameter log-normal distribution are given by:

$$\alpha_1 = \exp (m_n + \sigma_n^2 / 2) \quad (6.1)$$

$$\mu_2 = \alpha_1^2 (\exp (\sigma_n^2) - 1) \quad (6.2)$$

$$C_s = (\exp (\sigma_n^2) - 1)^{1/2} ((\exp (\sigma_n^2) + 2) \quad (6.3)$$

In Tables 6.1 - 6.3 these three coefficients, calculated from the expressions above with the help of maximum likelihood estimates of m_n and σ_n are compared to the observed ones. We can see that in general the expected value of the log-normal process is higher, but the differences are very small - one or two percent. The differences

TABLE 6.1 Observed and calculated monthly expected values for
96-1008 N Forsmöllan

Month	Observed	Calculated with parameters of	
		log-normal distribution	gamma distribution
Jan	15.0	15.4	15.0
Feb	16.0	16.6	16.0
March	14.0	14.3	14.0
April	15.0	15.2	15.0
May	8.0	8.3	8.0
June	5.0	5.0	5.0
July	5.0	5.0	5.0
August	5.0	4.9	5.0
Sept	6.0	6.0	6.0
Oct	8.0	8.1	8.0
Nov	11.0	11.0	11.0
Dec	14.0	14.3	14.0

TABLE 6.2 Observed and calculated monthly standard deviation for
96-1008 N Forsmöllan

Month	Observed	Calculated	
		log-normal distribution	gamma distribution
Jan	6.90	8.96	7.33
Feb	9.28	12.97	9.89
March	6.86	8.87	7.16
April	8.55	9.82	8.16
May	4.00	5.25	4.04
June	1.55	1.58	1.53
July	2.35	2.31	2.18
August	3.50	3.04	3.67
Sept	3.78	3.88	4.58
Oct	5.12	6.11	6.65
Nov	6.60	7.13	7.89
Dec	6.86	8.50	7.06

TABLE 6.3 Observed and calculated monthly skew coefficients for
96-1008 N Forsmöllen

Month	Observed	Calculated	
		log-normal distribution	gamma distribution
Jan	0.35	1.943	0.978
Feb	0.61	2.819	1.236
March	0.90	2.097	1.022
April	1.40	2.204	1.088
May	1.40	2.150	1.012
June	0.42	0.985	0.610
July	1.09	1.485	0.874
August	2.37	2.097	1.156
Sept	1.39	2.204	1.144
Oct	1.09	2.685	1.208
Nov	1.24	2.204	1.126
Dec	0.58	1.993	1.010

between the observed and calculated variances are larger. The values, calculated from the log-normal distribution, are always the largest except for some months, when they are approximately equal to the observed variances. We shall note that large differences are observed in connection with relatively high variation coefficient but, at the same time, low skew coefficient. The skew coefficient is generally larger for the log-normal distribution than for the observed values. Observed values of the skew coefficient are, as a rule, underestimated from small samples (see, for instance, Wallis et al 1974).

Let us look at the lag one correlation coefficient ρ . The mixed second order moment for the bivariate log-normal distribution can be calculated as:

$$E(x_1, x_2) = \exp \left(m_n^1 + m_n^2 + (\sigma_n^1)^2 / 2 + (\sigma_n^2)^2 / 2 + \right. \\ \left. + \rho_n \sigma_n^1 \sigma_n^2 \right)$$

where ρ_n is the correlation coefficient between the logarithms of flow modules. Using formula (4.16) and the above expressions for the expected value and the variance of the marginal log-normal distribution, the correlation coefficient is given by:

$$\rho(x_1, x_2) = (\exp(\sigma_n^1 \sigma_n^2 \rho_n) - 1) / ((\exp(\sigma_n^1)^2 - 1) / (\exp(\sigma_n^2)^2 - 1))^{1/2} \quad (6.4)$$

In Table 6.4 correlation coefficients, calculated from the original sample, are compared to the coefficients, calculated from the logarithm of flow and coefficients, calculated with the help of the above formula. We have already noted in (4.51), that the correlation coefficients of logarithms of flow are in general, larger than those of the original flow.

TABLE 6.4 Observed correlation coefficients of original and logarithm monthly runoff values and values calculated from eq (6.4) for 96-1008 N Forsmöllen

Month	Original runoff	Logarithm of runoff	Calculated from eq (6.4)
Jan	0.514	0.688	0.642
Feb	0.423	0.483	0.432
March	0.424	0.576	0.544
April	0.713	0.799	0.812
May	0.410	0.200	0.191
June	0.327	0.443	0.435
July	0.658	0.731	0.703
August	0.739	0.805	0.780
Sept	0.557	0.759	0.734
Oct	0.704	0.751	0.725
Nov	0.573	0.632	0.610
Dec	0.493	0.581	0.558

The correlation coefficients, calculated from eq (6.4), are usually also larger than those of the original flow, but they are always less than coefficients of the logarithms of flow.

We conclude, that when using the log-normal distribution, our model will have larger variability and will be higher correlated, than when using the normal distribution. The larger variability does not only give rise to larger coefficient of variation for certain simulated series but also larger deviation between parameters, calculated from different simulated series. This is illustrated in Table 6.5, where the range of estimated standard deviations of ten simulated series for 44 years from the normal and log-normal models, respectively, is given. As a comparison, a 99 % confidence interval for the case of normality is also shown.

To avoid the problem, that the moments are not preserved for the log-normal distribution, instead of using the maximum likelihood method and the correlation coefficients of the logarithms of flow, we can use eqs (6.1), (6.2) and (6.4) for parameter estimation.

Analytical expressions for the moments of the transformations eqs (3.22), (3.31 - 33) can not be found. Simulated series should be used to calculate them. The best results are, of course, achieved with the three parameter transformation, which gives values of the mean, variance and skew coefficient in accordance with the observed ones. The two parameter transformation $u = a x - b / x$ also preserves well these three parameters. In most cases the variance, calculated from simulated series, is somewhat higher than the one observed but not as high as for the lognormal model. The skewness is also lower than for the lognormal model. The other two two-parameter distributions preserve the variance and the skewness unsatisfactory. The transformation $u = a \sqrt{x} - b / x$ gives too high variance and skewness while the transformation $u = a x^2 - b / x$ preserves the variance well but gives small and in many cases negative skewness. The monthly correlation coefficients of the original sample are also not preserved, when using these transformations. Correlation coefficients calculated from simulated series, are usually smaller than those, calculated from transformations of the observed series, but larger than the original sample correlation coefficient.

TABLE 6.5 Range of standard deviation calculated from ten simulated series with a first order Markov model with normal and lognormal distributions, respectively for 96-1008 N Forsmöllan

Month	Distribution		99 % confidence interval for normal distribution
	Normal	Lognormal	
Jan	5.83 - 7.32	6.23 - 10.21	5.02 - 8.84
Feb	8.17 - 10.50	11.27 - 18.60	6.77 - 11.88
March	5.94 - 7.19	7.10 - 11.44	5.01 - 8.78
April	8.01 - 10.43	8.18 - 12.37	6.24 - 10.94
May	3.57 - 4.59	4.60 - 7.66	2.92 - 5.12
June	1.31 - 1.77	1.54 - 2.03	1.13 - 1.92
July	1.71 - 2.58	1.55 - 2.95	1.72 - 3.01
Aug	2.90 - 3.63	1.77 - 4.43	2.56 - 4.48
Sept	2.91 - 4.12	2.37 - 7.70	2.76 - 4.84
Oct	4.51 - 5.33	3.44 - 9.88	3.74 - 6.55
Nov	5.05 - 7.66	5.06 - 8.88	4.82 - 8.45
Dec	5.56 - 8.94	6.64 - 11.51	5.01 - 8.78

The theoretical expected value α_1 , variance μ_2 and skewness C_S of the gamma distribution are expressed as:

$$\alpha_1 = \gamma \beta \quad (6.4)$$

$$\mu_2 = \gamma \beta^2 \quad (6.5)$$

$$C_S = 2 / \sqrt{\gamma} \quad (6.6)$$

In tables 6.1 - 6.3 values of these three coefficients are shown with γ and β derived by the maximum likelihood method. The standard deviations calculated in this way are generally larger than the observed ones, but not as large as those for the lognormal distribution. The skewness is usually also larger than the observed values.

With a correction for biases in the observed skew coefficients the calculated values still seem to be small.

Let us now turn to the bivariate gamma distribution eq (4.7). To calculate the mixed second order moment of this distribution we shall use the following relation:

$$\begin{aligned} (\Gamma(\gamma) \beta^{-\gamma})^{-1} \int_0^{\infty} x^{n+\gamma-1} \exp(-x/\beta) dx &= \gamma(\gamma+1) \dots \\ &\dots (\gamma+n-1) \beta^n \end{aligned}$$

The mixed second order moment between the variables X_1 and X_2 is now expressed by:

$$\begin{aligned} E(X_1, X_2) &= \gamma_1 \gamma_2 \beta_1 \beta_2 + c_1 (g_1^1 \gamma_1 \beta_1 - g_2^1 \gamma_1 \\ &(\gamma_1 + 1) \beta_1) \cdot \\ &(h_1^1 \gamma_2 \beta_2 - h_2^1 \gamma_2 (\gamma_2 + 1) \beta_2) + c_2 (g_1^2 \gamma_1 \beta_1 - \\ &g_2^2 \gamma_1 (\gamma_1 + 1) \beta_1 + g_3^2 \gamma_1 (\gamma_1 + 1) (\gamma_1 + 2) \beta_1) (h_1^2 \gamma_2 \beta_2 - \\ &- h_2^2 \gamma_2 (\gamma_2 + 1) \beta_2 + h_3^2 \gamma_2 (\gamma_2 + 1) (\gamma_2 + 2) \beta_2) + \dots \end{aligned}$$

where g_j^i and h_j^i , $i = 1, 2, \dots$; $j = 1, 2, \dots, i+1$ are the coefficients in the Laguerre polynomials. By inserting their expressions in the above equation, it can be simplified to:

$$E(X_1, X_2) = \gamma_1 \gamma_2 \beta_1 \beta_2 + c_1 \sqrt{\gamma_1 \gamma_2} \beta_1 \beta_2$$

We can now finally find the correlation coefficient for the bivariate gamma distribution by using eqs (6.4) and (6.5):

$$\begin{aligned} \rho(X_1, X_2) &= (\gamma_1 \gamma_2 \beta_1 \beta_2 + c_1 \sqrt{\gamma_1 \gamma_2} \beta_1 \beta_2 - \\ &- \gamma_1 \gamma_2 \beta_1 \beta_2) / \gamma_1 \gamma_2 \beta_1 \beta_2 = c_1 \end{aligned}$$

In (4.5.3) we have already found that the parameter c_1 is identical with the correlation coefficient. The bivariate gamma distribution will, thus, preserve this coefficient.

6.2 Moments of annual runoff

The fact, that a model preserves the moments of monthly runoff, does not guarantee that the moments of the annual runoff, as a sum of twelve monthly runoff values, are preserved. We shall also draw attention to the fact that the moments of annual runoff are dependent on how the year is defined. This was first pointed out in an article by Svanidze et al, 1970, no explanation to the phenomena was given, however. Differences in mean values are very small and the only cause is that the monthly start and end values of the series are different when changing the division into years. In Tables 6.6 - 6.8 are shown calculated values of variation coefficients, skew coefficients and first order correlation coefficients as a function of the beginning month of the year for some of the observation series in the study. In these three coefficients we notice a small but clear periodicity. The variations in the skew coefficients are the largest. We can further observe that the variation coefficient for the Northern groups of rivers has its maximum value usually for June as the beginning of the year and at the same time the correlation coefficient has its minimum.

TABLE 6.6 Annual variation coefficients as a function of the beginning month of the year

Observation point	Month											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1-589	0.19	0.19	0.19	0.19	0.19	0.23	0.21	0.19	0.18	0.18	0.18	0.19
4-10	0.13	0.13	0.13	0.13	0.12	0.14	0.15	0.13	0.12	0.11	0.11	0.12
28-444	0.18	0.18	0.18	0.18	0.18	0.19	0.18	0.18	0.19	0.18	0.18	0.18
38-71	0.18	0.18	0.18	0.18	0.18	0.22	0.22	0.21	0.20	0.18	0.18	0.18
74-177	0.33	0.33	0.34	0.34	0.35	0.36	0.36	0.37	0.36	0.36	0.35	0.33
96-1008	0.25	0.26	0.28	0.27	0.26	0.26	0.27	0.27	0.27	0.27	0.27	0.26
67-154	0.22	0.22	0.22	0.23	0.23	0.24	0.24	0.24	0.23	0.23	0.22	0.22

TABLE 6.7 Annual skew coefficients as a function of the beginning month of the year

Observation point	Month											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1-589	0.71	0.69	0.68	0.67	0.59	0.50	0.47	0.31	0.30	0.43	0.63	0.71
4-10	-0.21	-0.22	-0.22	-0.22	-0.21	0.09	-0.01	-0.02	-0.09	0.02	-0.22	-0.29
28-444	0.45	0.43	0.41	0.41	0.40	0.31	0.29	0.31	0.36	0.45	0.48	0.45
38-71	0.16	0.18	0.22	0.24	0.22	0.64	0.16	-0.07	-0.18	-0.30	-0.00	0.20
74-177	0.24	0.29	0.34	0.28	0.27	0.27	0.32	0.35	0.36	0.34	0.37	0.28
96-1008	0.22	0.30	0.07	0.07	0.15	0.24	0.27	0.19	0.20	0.26	0.35	0.30
67-154	0.32	0.29	0.28	0.32	0.41	0.48	0.52	0.53	0.52	0.51	0.50	0.51

TABLE 6.8 Annual lag one correlation coefficients as a function of the beginning month of the year

Observation point	Month											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1-589	0.17	0.17	0.17	0.17	0.16	0.02	0.10	0.19	0.32	0.34	0.27	0.21
4-10	0.09	0.09	0.09	0.09	0.10	-0.02	-0.05	-0.08	0.03	0.07	0.10	0.08
28-444	0.16	0.15	0.15	0.15	0.15	0.05	0.14	0.13	0.10	0.13	0.15	0.14
38-71	0.16	0.16	0.14	0.13	0.11	-0.06	-0.07	-0.10	0.05	0.17	0.16	0.14
74-177	0.24	0.29	0.34	0.28	0.27	0.27	0.32	0.35	0.36	0.34	0.37	0.28
96-1008	0.16	0.06	-0.03	-0.02	-0.03	-0.07	-0.10	-0.08	-0.06	-0.01	0.13	0.19
67-154	0.39	0.40	0.40	0.37	0.32	0.27	0.26	0.25	0.25	0.27	0.29	0.31

We derive annual runoff X as the sum of twelve monthly values:

$$X = (X_1 + X_2 + \dots + X_{12}) / 12 \quad (6.9)$$

The expected value of annual runoff is thus:

$$E(X) = (E(X_1) + E(X_2) + \dots + E(X_{12})) / 12 = m \quad (6.10)$$

It is obvious at once that the expected value is not dependent on the definition of the year.

The expression for the annual variance can be written as:

$$\begin{aligned} E((X - m)^2) &= E((X_1 - m_1) + (X_2 - m_2) + \dots + (X_{12} - m_{12}))^2 \\ &/ 144 = E((X_1 - m_1)^2 + (X_2 - m_2)^2 + \dots + (X_{12} - m_{12})^2 + \\ &+ 2(X_1 - m_1)(X_2 - m_2) + 2(X_1 - m_1)(X_3 - m_3) + \dots + \\ &+ 2(X_1 - m_1)(X_{12} - m_{12}) + 2(X_2 - m_2)(X_3 - m_3) + \dots + \\ &+ 2(X_{11} - m_{11})(X_{12} - m_{12})) / 144 = \\ &= \left(\sum_{i=1}^{12} \text{Var}(X_i) + 2(\text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \dots + \right. \\ &+ \text{Cov}(X_1, X_{12}) + \text{Cov}(X_2, X_3) + \dots + \text{Cov}(X_{11}, X_{12})) \bigg) / 144 \end{aligned} \quad (6.11)$$

From this expression we can see that the annual variance is dependent on the marginal monthly distributions, as well as the correlation structure within the year. When the covariance matrix for the monthly runoff values is known, we can calculate the annual variance. Of course, it can be also directly calculated from the observed annual runoff values. There is a small difference in the variance calculated in these two different ways. It is caused, perhaps, by errors in the estimated parameters. For example, for the series 96-1008 N Forsmöllen the variance, calculated from annual values, is 6,5, while only 5,4 is got, when the calculation is performed from monthly values with the

help of equation (6.11). If we now define the year from February to January, the annual variance, calculated from monthly runoff values is 5,7. The fact, that it differs from the above figure, calculated in the same way, but with another definition of the year, is explained by the instationarity of the runoff process. In the latter case the covariances between X_j and X_i ; $i = 1, \dots, 12$ are changed for the covariances between X_i ; $i = 1, \dots, 12$ and X_j and they are not equal. We can continue with the calculation of the annual variances for different beginning months of the year with the help of equation (6.11) and find the same pattern of variation as the one, given in Table 6.6. The differences are in themselves not significant. This effect can be, of course, simulated in a model only in case the nonstationarity in the marginal moments as well as in the correlation coefficients are considered.

In the models described above this instationarity is considered. The value of the variance changes, however with the type of marginal distribution, the order of the Markov process and the individual values of the correlation coefficients. For the example discussed above the annual variance, (Jan-Dec), calculated from 440 years of simulated flow with normal marginal distributions, is 7,68. For the log-normal distribution we find in the same way the value 12,2. By analogy with the marginal monthly variance the annual variance is higher when we use transformations of the normal distribution than when we use the normal distribution itself. This is not only caused by the fact that monthly variances are higher but by the higher correlation coefficients for the transformed variables. The relative differences are usually not as large as in the examples above. For the series 1-509 Kallio the observed annual variance is 93,5. From 610 years of simulated annual values with a first order normal model we find the value 88,6, while the lognormal model gives 99,7. For a second order log-normal model the annual variance 75,0 was found and the same value was obtained for a third order model.

For the annual skew coefficient we can derive a formula by analogy with the one for the variance eq (6.11). It will contain skew coefficients for the individual months as well as mixed third order moments. Annual skew coefficients, however, were calculated only from simulated series. Individual simulated series, with the same length as the

observed ones, gave the values like those got from the observed series. The deviation between the different simulations was usually large. Almost all of the observed skew coefficients have relatively small positive values. The normal model gave small skewness but with both positive and negative sign. The lognormal model gave small positive value but also relatively large positive skewnesses, while, for instance, the transformation $= a x^2 - b / x$ gave small positive skewnesses but also large skewnesses that were negative.

The annual distribution of the flow is, thus, not exactly preserved for any of the models used. To find a model, that actually does this we should work with the marginal monthly distributions as well as the order of model. The normal model preserves the annual variance best and also gives small standard deviation in this parameter between different simulated series. The lognormal and the transformation eqs (3.22) and (3.29) models preserves the annual skewness the best.

A still more complicated question is to determine the type of the annual process, when it is built up by the sum of the monthly processes. Is the sum of twelve monthly processes actually a Markov process itself? We think, that in general, this question is answered with no. This problem should be studied further. Here we shall only discuss the first order correlation of annual runoff

We have observed that the annual correlation coefficient is dependent on the definition of the year. It is obvious, that the monthly correlation coefficients near the beginning and the end of the year will have the main influence on it. If the beginning of the year is in a period with low correlation, say snowmelt, the annual correlation will be lower than if the beginning is in a period with high corre-

lation. For the generation of annual flow an independent model can, thus, be used with a proper definition of the year.

The observed annual correlation coefficients have usually small positive values. Most of them do not differ significantly from zero. The annual correlation coefficients, calculated from simulated series, show much larger deviations between various simulations than the deviations in the values, observed in different points. In Table 6.9 is shown the range of annual correlation coefficients, calculated from ten simulated series of 61 years from 1-589 Kallio with some different orders of the monthly generating models and some different distributions. We see relatively large negative values of the correlation coefficients. A comparison with the observed annual correlation coefficients from various points does not make it probable that they are derived from the same kind of generating process. The conclusion is, thus, that the models cannot preserve the annual correlation coefficient well, although correlation coefficients, calculated from some individual series, coincide well with the observed ones.

TABLE 6.9 Range of annual (Jan-Dec) correlation coefficients, calculated from ten simulated series of different models for 1-589 Kallio

Type of model	Range of annual corr coeff
First order normal	- 0.21 - 0.13
First order with	
trf $u = a x^2 - b / x$	- 0.44 - 0.29
First order lognormal	- 0.15 - 0.23
Second order lognormal	- 0.29 - 0.19
Third order lognormal	- 0.20 - 0.16

6.3 Other properties

There are several other properties of importance that should be preserved by the models. Namely - extreme events and crossing properties. A visual investigation of simulated series can be a first rough method to see to which extent a model possesses them. In Figs 6.1 and 6.2 two examples of simulated series with different first order models are shown. The corresponding observed series are shown in Figs 3.4 and 3.6. The distributions used are the lognormal and normal distributions and the three parameter transformation of the unit normal distribution. The same set of random numbers is used for the three series. We can see at once that series from the lognormal and the three parameter transformation differ very little. Due to the fact, that the normal distribution also gives negative values, the series, got from this distribution, are characteristically different from the two other simulated series and also from the observed series. The normal distribution has no skewness and therefore, does not give the same high values as the other series. It is obvious, that the normal model can preserve neither extreme events nor crossing properties of the observed series. The other two models are more likely to be able to do this.

Crossing properties as the number of upcrossings and downcrossings, the mean run length and mean run sums of upcrossings and downcrossings were calculated for simulated series from 1-589 Kallio and 96-1008 N Forsmöllan. Different orders of the models were tested with the normal and lognormal distributions. Ten simulations of the length, equal to that of the observed series, were used. It was found, that the order of the model had no significant influence on the crossing properties. The lognormal distribution gave results, that were more like the observed series than those, given by the normal one. A five months period for 1-589 Kallio with very low flow could not, of course, be described by any model.

6.4 General conclusions

Characteristics of the observed series should be within the confidence limits of the corresponding characteristics of the model for series of equal length. It seems, that such criteria for the acceptability of a model is more proper than the condition that the true properties of a model, i.e. properties derived from theoretical calculations or very



Fig 6.1 Examples of simulated series with different models.

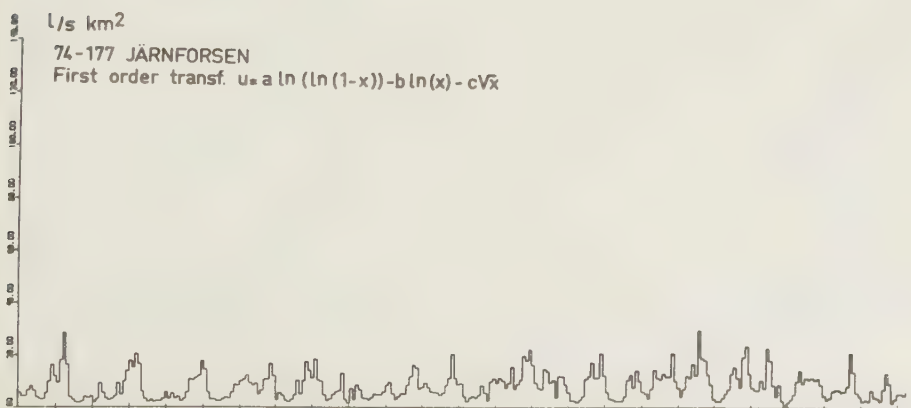
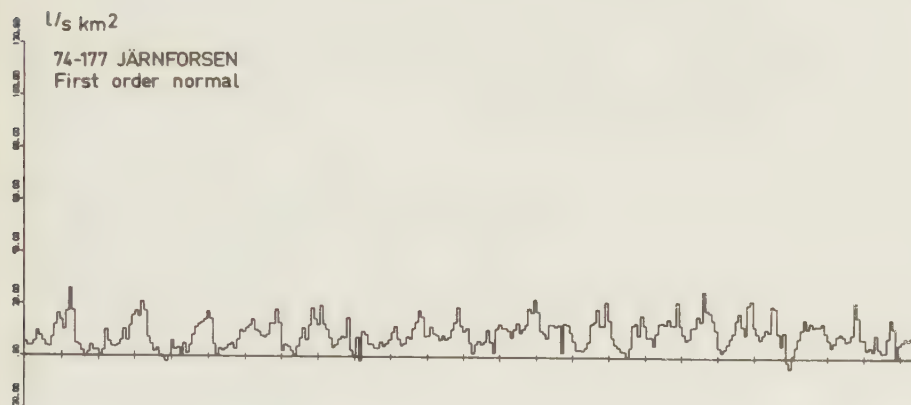
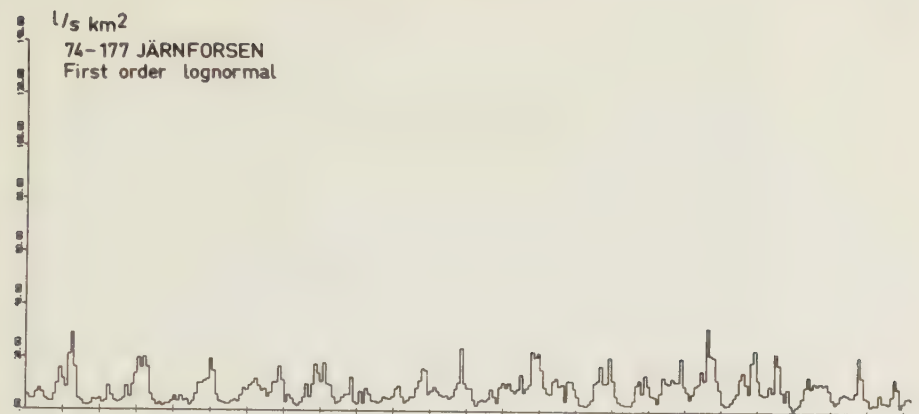


Fig 6.2 Examples of simulated series with different models.

long simulated series, are exactly the same as those for the observed series. However, confidence limits are usually difficult to find theoretically except for the normal law case. Monte-Carlo technique should be used to derive them. With this approach, the hypothesis, that the observed series are realizations from any of the models, can be tested. This kind of tests does not seem very strong and we are not sure to select one proper order of the model or one proper distribution. Several models can give traces with properties like the observed ones in respect to monthly and annual moments and crossing properties. In Fig 6.3 is shown an example of a confidence interval of one standard deviation around the expected monthly distribution of runoff of 30-years long series of the outflow from lake Bolmen, (98-1048), generated from a first order lognormal model, together with the observed distributions. The fact, that the observed distributions to a small extent lie outside this interval, is admissible in the theory for confidence intervals.

We have now discussed the problem taking as a basis the series at only one point. It was commented above, that for instance, the annual correlation and the annual skewness calculated from simulated series obtain values, that may be in agreement with the observed ones, but also such values, that one can doubt them to be observed in reality. This is especially the case when we compare to the values, observed for a large territory. The material in this study is too small to draw any general conclusions in this respect and further studies are needed. To improve the accuracy in the structure it is also necessary to use more completely the information about the physics of the runoff process. We believe, for instance, that better ideas of the correlation structure can be gained in this way.

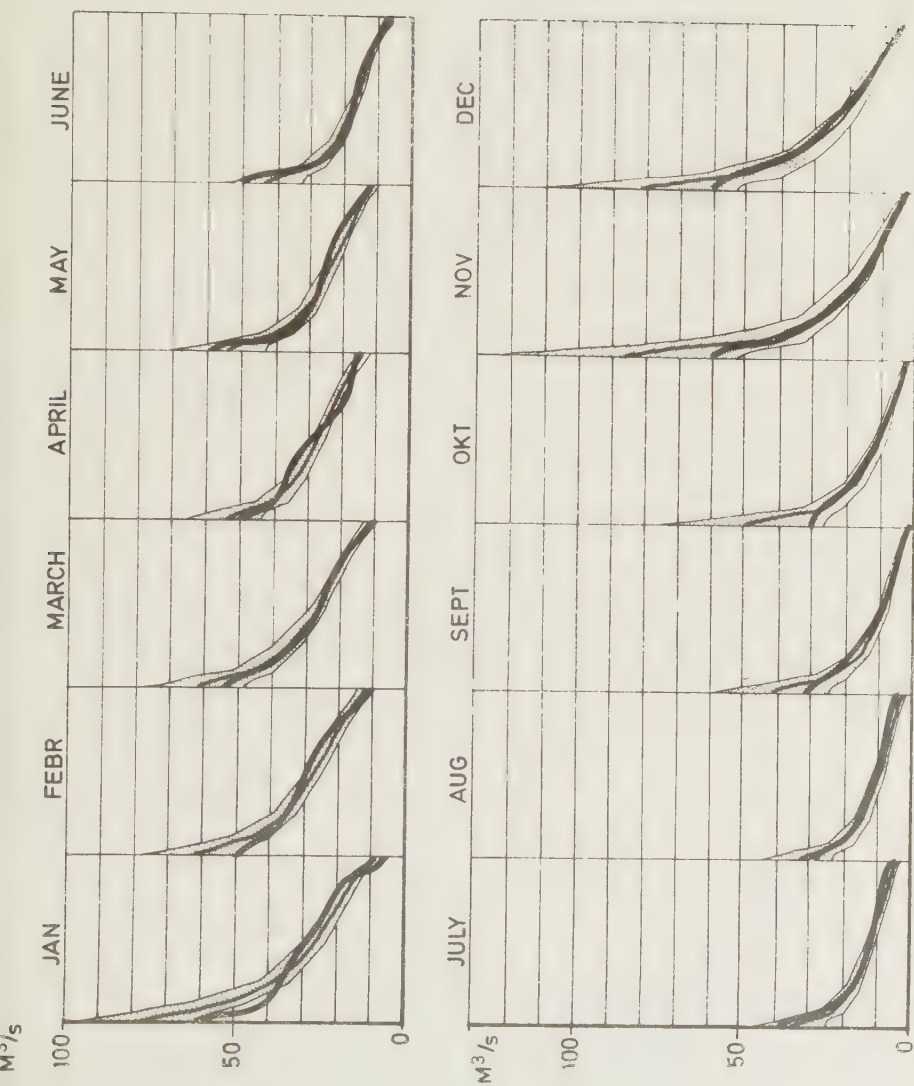


Fig 6.3 Observed and calculated, with confidence intervals, distributions of monthly outflow from the lake Bolmen.

7. CONCLUSIONS

The models for monthly river runoff for Swedish conditions, described in the present work can be, first of all, understood as a working tool that can be applied to solve certain water management problems. They have a wide range of applications. Some properties of the natural runoff process are not certainly fully preserved, and the relevance of these properties for a given design or operation of a water management installation should be analysed before the model is applied. However, at present these questions are not investigated enough. It can be advisable, therefore, to check the sensitivity of design and operation to the changes in distribution functions, orders of Markov process and parameter values.

The structure of the model, i.e. the analytical form of the distribution functions and the order of the Markov process, is to the most extent determined from the observational data, as well as the parameters of the model. Physical interpretation of parameters can be done, especially for the correlation coefficients. The uncertainty in the choice of the structure can be decreased by its testing on a homogenous region.

It has been shown that the structure is not very sensitive to the differences in the physiographic conditions. These differences are only reflected by parameter values. This on one hand shows that the model is very flexible but on the other hand, in our opinion, the structure should reflect the specific physical conditions. To gain this a more physically based approach for the construction of marginal and multivariate distributions, than the one shown here should be used. The theoretical difficulties, however, are great. Meanwhile we use the "black box" approach but with a regional analysis.

REFERENCES

- Aitchison, J and Brown, J A C, The log-normal distribution, Cambridge University Press, London, 1957.
- Alekseev, G A, Grafoanaliticheskie sposoby opredelenija i privedenija k dlitel'nomu periodu nabludenij parametrov krivykh raspredelenija, Trudy GGI, Vyp 73, Leningrad, 1960.
- Alekseev, G A, K voprosu opredelenija empiricheskikh kvantilej i koeffitsienta korrelyatsii, Meteorologiya i Gidrologiya, No 4, 1963.
- Alekseev, G A, Obektivnye metody vyravnivaniya i normalisatsii korrelyatsionnykh svyazej (S primerami iz gidrometeorologii), Gidrometeoizdat, Leningrad, 1971.
- Andersson, R L, Distribution of the serial correlation coefficients, Annals of Math Statistics, Vol 8, No 1, 1941.
- Askew, A J; Yeh, W W-G and Hall, W A, A comparative study of critical drought simulation, Water Resources Research, Vol 7, No 1, 1971.
- Bartlett, M S, Some aspects of the time-correlation problem in regard to tests of significance, Journal of the Royal Statistical Society, Vol XCVIII, 1935.
- Beard, L R, Use of interrelated records to simulate streamflow, Journal of the Hydraulic Division, Proc ASCE, Vol 91, No HY 5, 1965.
- Benson, M A, Uniform flood frequency estimating methods for federal agencies, Water Resources Research, Vol 4, No 5, 1968.
- Blochinov, E G, Otsenka parametrov sluchajnykh kolebanij rechnogo stoka metodom naibolshego pravdopodobiya, Trudy Gidroproekta, No 12, 1964.
- Blochinov, E G and Sarmanov, O V, Gammakorrelyatsiya i ee ispolzovanie pri raschetakh mnogoletnego regulirovaniya rechnogo stoka, Trudy GGI, Vyp 143, Leningrad, 1968.
- Blom, G, Statistical estimates and transformed β -variables, Almqvist & Wiksell, Stockholm, 1958.
- Borgman, L E and Amoroch, J, Some statistical problems in hydrology, Intern Statist Inst Review, Vol 38, No 1, 1970.
- Box, G E P and Jenkins, G M, Time series analysis forecasting and control, Holden Day, San Fransisco, 1970.
- Brownlee, K A, Statistical theory and methodology in science and engineering, Second edition, Third printing, J Wiley & Sons, New York, 1967.

- Burges, S J and Linsely, R K, Some factors influencing required reservoir storage, *Journal of the Hydraulic Division, Proc ASCE*, Vol 97, No Hy 7, 1971.
- Burges, S J, Some problems with log-normal Markov runoff models, *Journal of the Hydraulic Division, Proc ASCE*, Vol 98, No Hy 9, 1972.
- Chow, V T, Statistical and probability analysis of hydrologic data. Part 1: frequency analysis, in "Handbook of applied hydrology", Mc Graw-Hill, New York, 1964.
- Cohen, A C, Estimating parameters of logarithmic normal distributions by maximum likelihood, *Journal of American Statist Assoc*, Vol 46, p 206, 1951.
- Cramer, H, Mathematical methods of statistics, Princeton University Press, Princeton, 1948 (11-th printing 1966).
- Crosby, D S and Maddock, T III, Estimating coefficients of a flow generator for monotone samples of data, *Water Resources Research*, Vol 6, No 4, 1970.
- Dalrymple, T, Flood-frequency analysis manual of hydrology: Part 3. Flood flow techniques, U S Geological Survey Water Supply Paper, 1543 A, 1960.
- Durbin, J, Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica*, Vol 38, p 410, 1970.
- Durbin, J and Watson, G S, Testing for serial correlation in least squares regression, *Biometrika*, Vol 37, p 409, 1971.
- Fiering, M B, Queuing theory and simulation in reservoir design, *Journal of the Hydraulic Division, Proc ASCE*, Vol 87, No Hy 6, 1961.
- Fiering, M B, Sequential generation of rainfall and runoff data, *Journal of the Hydraulic Division, Proc ASCE*, Vol 90, No HY 5, 1964.
- Fröberg, C-E, *Lärobok i numerisk analys*, Scandinavian University Books, Stockholm, 1962.
- Gottschalk, L and Falk, J, Samband mellan avrinningsparametrar och fysiografiska karakteristika, en inledande studie för Skåne och Sjöland, Opublicerat material från institutionen för vattenbyggnad, LTH, Lund, 1973.
- Gumbel, E J, *Statistics of extremes*, Columbia University Press, New York, 1958.

- Hasting, S C, Approximations for digital computers, Princeton University Press, Princeton, 1955.
- Jenkins, G M and Watts, D G, Spectral analysis and its applications, Holden-Day, San Fransisco, 1968.
- Johnson, N L, Systems of frequency curves generated by methods of translation, *Biometrika*, Vol 36, p 149, 1949a.
- Johnson, N L, Bivariate distributions based on simple translation systems, *Biometrika*, Vol 36, p 297, 1949b.
- Kaczmarek, Z, *Metody statystyczne w hydrologii i meteorologii*, Warszawa, 1970.
- Kapitanova, N P, Metodika rajonirovaniya statisticheskikh kharakteristik vnutrigodovogo raspredeleniya stoka, *Problemy Gidroenergetiki i vodnogo khozyajstva*, Vyp 8, Alma Ata, 1971.
- Kartvelishvili, N A, Teoriya veroyatnostnykh protsessov v gidrologii i regulirovanii rechnogo stoka, *Gidrometeoizd*, Leningrad, 1967.
- Kartvelishvili, N A, Regulirovanie rechnogo stoka, *Gidrometeoizd*, Leningrad, 1970.
- Kartvelishvili, N A and Korganova, N S, Integralnye uravneniya regulirovaniya rechnogo stoka, *Problemy gidroenergetiki i vodnogo khozyajstva*, Vyn 10, Alma Ata, 1973.
- Kendall, M G, Note on bias in the estimation of autocorrelation, *Biometrika*, Vol 41, p 403, 1954.
- Kendall, M G, *Time series*, Griffin, London 1973.
- Kiktenko, V A and Oshirov, E N, Metodika analiticheskogo rascheta mnogoletnego regulirovaniya stoka pri bolshoy emkosti vodokhranilisha, *Problemy gidroenergetiki i vodnogo khozyajstva*, Vyp 2, Alma Ata, 1964a.
- Kiktenko, V A and Oshirov, E N, Analiticheskij metod opredeleniya otdach i sbrosov pri mnogoletnem regulirovanii stoka, *Problemy gidroenergetiki i vodnogo khozyajstva*, Vyp 2, Alma Ata, 1964b.
- Kiktenko, V A and Oshirov, E N, Razvitie obobshennogo metoda raschetov mnogoletnego regulirovaniya stoka na osnove integralnykh uravnenii, *Problemy gidroenergetiki i vodnogo khozyajstva*, Vyp 3, Alma Ata, 1965.
- Klemés, V, Comments on "Adequacy of markovian models with cyclic componen components for stochastic streamflow simulation" by I Rodriguez-Iturbe, D R Dawdy and L E Garcia, *Water Resources Research*, Vol 8, No 6, 1972.
- Korganova, N S, Ob analiticheskikh vypazhenyakh raspredelenij veroyatnostej gidrologicheskikh velichin, Personal Communication, 1971.

- Kottegoda, N T, Effect of skewness in three stochastic pentad river flow models on crossing properties of synthesized data, Water Resources Research, Vol 10, No 3, 1974.
- Kritskij, S N and Menkel, M F, O priemakh issledovaniya sluchajnykh kolebanij rechnogo stoka, sbornik "Stok i gidrologicheskie raschety", Ser IV, Vyp 29, Gidrometeoizdat, 1946.
- Kritskij, S N and Menkel, M F, O primeneniij metoda nailbol'shego pravdopodobiya k vyborochnoj otsenke statisticheskikh parametrov rechnogo stoka, Izvestiya A N SSSR, Otd tekhn nauk, No 4, 1949.
- Kritskij, S N and Menkel, M F, Gidrologicheskie osnovy rechnoj gidrotehniki, Izdatel'stvo A N SSSR, 1950.
- Kritskij, S N and Menkel, M F, O primeneniij metoda statisticheskikh ispytaniy k issledovaniyu kolebanij rechnogo stoka i k raschetam rezhima ispol'zuyushikh ego ustanovok, Problemy gidroenergetiki i vodnogo khozyajstva, Vyp 2, Alma Ata, 1964.
- Kupriyanov, V V, Raskhod i isparenije s basejnov skandinavskogo poluostrova, Trudy GGI, Vyp 78, 1960.
- Lancaster, N O, The structure of bivariate distributions, Annals of Math Statistics, Vol 29, p 719, 1958.
- Lyttkens, E, Standard errors of regression coefficients in the case of autocorrelated residuals, chapter 3 in "Proceedings of the symposium of time series analysis" ed Rosenblatt, M, John Wiley & Sons, New York, 1963.
- Mandelbrot, B B and Wallis, J R, Noah, Joseph, and operational hydrology, Water Resources Research, Vol 4, No 5, 1968.
- Mandelbrot, B B and Wallis, J R, Computer experiments with fractional gaussian noises.
- Part 1. Averages and Variances.
- Part 2. Rescaled ranges and spectra.
- Part 3. Mathematical appendix.
- Water Resources Research, Vol 5, No 1, 1969 a,b,c.
- Mandelbrot, B B and Wallis, J R, Some long run properties of geophysical records, Water Resources Research, Vol 5, No 2, 1969.
- Mc Mahon, T A; Godner, G P and Phillips, C, Singel and multi-site operational hydrology, Nordic Hydrology, Vol 3, No 3, 1972.
- Markovic, R D, Probability functions of best fit to distributions of annual precipitation and runoff, Colorado State University Hydrology paper, No 8, Fort Collins, Colorado, 1965.
- Marriott, F H C and Pope, J A, Bias in the estimation of autocorrelations, Biometrika, Vol 41, p 390, 1954.

- Matalas, N C, Mathematical assessment of synthetic hydrology, *Water Resources Research*, Vol 3, No 4, 1967.
- Matalas, N C and Huzzzen, C S, A property of the range of partial sums, *International Hydrology symposium*, Fort Collins, 1967.
- Matalas, N C and Wallis, J R, Statistical properties of multivariate fractional noise processes, *Water Resources Research*, Vol 7, No 6, 1971.
- Meija, J M, On the generation of multivariate sequences exhibiting the thirst phenomena and some flood frequency analysis, Ph D dissertation, Colorado State University, Fort Collins, Colorado 1971.
- Melin, R, Vattenföringen i Sveriges floder, SMHI, Meddelanden Serie D, No 6, Stockholm 1955.
- Melin, R, Hydrologi i Norden, Utbildningsförlaget, 1970.
- Moran, P A P, Statistical inference with bivariate gamma distributions, *Biometrika*, Vol 63, No 3, p 627, 1969.
- Mustonen, S E, Effects of climatological and basin characteristics on annual runoff, *Water Resources Research*, Vol 3, No 1, 1967.
- Nordseth, K, Morfometrisk kontroll av avrinningsregimet i Norske vassdrag. En förelöpig og generell redegjørelse, Femte Nordiska IHD-Feltsymposium, NTH, Trondheim 22-24 Maj, 1973.
- O'Connell, P E, A simple stochastic modelling of Hursts law, *International symposium on mathematical models in hydrology*, Vol 1, Part 1, Warszawa, 26-31 July, 1971.
- Person, K, On the mean character and variance of ranked individuals and on the mean and variance of the intervals between ranked individuals, *Biometrika*, Vol 23, p 364, 1931. *Biometrika*, Vol 24, p 203, 1932.
- Quenouille, M H, A large-sample test for the godness of fit of autoregressive schemes, *Journal of the Royal Statistical Society*, Vol CX, p 123, 1948.
- Quenouille, M H, Notes on bias in estimation, *Biometrika*, Vol 43, p 353, 1956.
- Quimpo, R G, Stochastic models of daily river flow sequences, Colorado State University, Hydrology paper, No 18, Fort Collins, Colorado, 1967.
- Ratkovich, D YA, Zakonomernost cheredovaniya malovodnykh i mnogovodnykh let kak osnova raschetov regulirovaniya rechnogo stoka, *Trudy GGI*, Vyp 143, Leningrad, 1968a.
- Trudy GGI, Vyp 160, Leningrad, 1968b.

- Trudy GGI, Vyp 162, Leningrad, 1969.
- Trudy GGI, Vyp 180, Leningrad, 1970.
- Reznikovskij, A Sh, ed, Vodnoenergeticheskie razchety metodom Monte-Carlo, Energiya, Moskva, 1969.
- Rodrigues-Iturbe, I; Dawdy, D R and Garzia, L E, Adequacy of markovian models with cyclic components for stochastic streamflow simulation, Water Resources Research, Vol 5, No 5, 1971.
- Rodrigues-Iturbe, I; MeiJa, J M and Dawdy, D R, Streamflow simulation: 1. A new look at markovian models, fractional gaussian noise and crossing theory, Water Resources Research, Vol 8, No 4, 1972.
- Roesner, L A and Yevjevich, W M, Mathematical models for time series of monthly precipitation and monthly runoff, Colorado State University, Hydrology paper, No 15, Fort Collins, Colorado, 1966.
- Salas-La Cruz, J D and Boes, D C, Expected range and adjusted range of hydrologic sequences, Water Resources Research, Vol 10, No 3, 1974.
- Sarmanov, O V, Psevdonormalnaya korrelyatsiya i razlichnye ee obobsheniya, Doklady A N, SSSR, Tom 132, No 2, 1960.
- Sarmanov, I O, Obobshennaya simmetrichnaya gamma-korrelyatsiya, Doklady A N, SSSR, Tom 179, No 6, 1968.
- Sarmanov, I O, Nesimmetrichnaya gamma-korrelyatsiya, Trudy GGI, Vyp 162, Leningrad, 1969.
- Smirnov, V I, Kurs vyshej matematiki, Tom III, Chast I, izdanie vosmoe, Gos Izd Fis-Mat-Literatury, Moskva, 1958.
- Stidd, C K, Cube-root-normal precipitation distributions, Transactions American Geophysical Union, Vol 34, No 1, 1953.
- Stidd, C K, The n :th root normal distribution of precipitation, Water Resources Research, Vol 6, No 4, 1970.
- Svanidze, G G, Metodika stokhasticheskogo modelirovaniya gidrologicheskikh ryadov i nekotorye voprosy mnogoletnego regulirovaniya rechnogo stoka, Trudy Inst Energetiki A N, GSSR, Tom XIV, Tbilisi, 1961.
- Svanidze, G G, Osnovy rascheta regulirovaniya rechnogo stoka metodom Monte-Carlo, Izdatelstvo Metsniereba, Tbilisi, 1964.
- Svanidze, G G; Piranashvili, Z A and Kilasoniya, A N, K voprosu ob izmenenii statisticheskikh parametrov rechnogo stoka pri razlichnoj razrezke gidrografa na uchastki s godovymi intervalami osredneniya, Meteorologiya i Gidrologiya, No 6, 1970.
- Thom, H C S, A note on the gamma distribution, Monthly Weather Review, Vol 86, No 4, 1958.

- Tollan, A, Variability of seasonal and annual runoff in Norway, Nordic Hydrology, Vol 3, No 2, 1972.
- Valencia, D R and Schaake, J C, Disaggregation processes in stochastic hydrology, Water Resources Research, Vol 9, No 3, 1973.
- Ventsel, E S, Teoriya veroyatnostej, Izd Nauka, Moskva, 1964.
- Vere-Jones, D, The infinite divisibility of a bivariate gamma distribution, Sankhya, A 29, p 421, 1967.
- Wallis, J R and Matalas, N C, Correlogram analysis revisited, Water Resources Research, Vol 7, No 6, 1971.
- Wallis, J R and O'Connell, P E, Small sample estimation of ρ_1 , Water Resources Research, Vol 8, No 3, 1972.
- Wallis, J R and Matalas, N C, Sensitivity of reservoir design to the generating mechanism of inflows, Water Resources Research, Vol 8, No 3, 1972.
- Wallis, J R; Matalas, N C and Slack, J R, Just a moment! Water Resources Research, Vol 10, No 2, 1974.
- Wold, H, On least square regression with autocorrelated variables and residuals, Bulletin de l'institut international de statistique, Berne, 1950.
- Weibull, W, Fatigue testing and analysis of results, Pergamon Press, 1961.
- Yevjevich, V, Fluctuations of wet and dry years 2. Analysis by serial correlation, Colorado State University Hydrology paper, No 4, Fort Collins, Colorado, 1964.
- Yevjevich, V, Structural analysis of hydrologic time series, Colorado State University Hydrology papers, No 56, Fort Collins, Colorado, 1972a.
- Yevjevich, V, Probability and statistics in hydrology. Water Resources Publications, Fort Collins, Colorado, 1972b.
- Young, G K and Pisano, W C, Operational hydrology using residuals, Journal of the Hydraulic division. Proc ASCE, Vol 94, No Hy4, 1968.
- Åström, K J; Bohlin, T and Wensmark, S, Automatic control of linear stochastic dynamic models for stationary industrial processes with random disturbances using normal operating records, IBM Nordic Laboratory Sweden, Report TP 18.150, 1965.

